

UNIVERSIDADE DE LISBOA

Faculdade de Ciências

Departamento de Informática



DESENVOLVIMENTO DE SISTEMA DE  
INFORMAÇÃO E APLICAÇÃO WEB PARA  
PERMEABILIDADE HEMATO-ENCEFÁLICA

Luís Carlos Mota Pinheiro

PROJECTO

MESTRADO EM ENGENHARIA INFORMÁTICA

Sistemas de Informação

2013



UNIVERSIDADE DE LISBOA

Faculdade de Ciências

Departamento de Informática



DESENVOLVIMENTO DE SISTEMA DE  
INFORMAÇÃO E APLICAÇÃO WEB PARA  
PERMEABILIDADE HEMATO-ENCEFÁLICA

Luís Carlos Mota Pinheiro

PROJECTO

MESTRADO EM ENGENHARIA INFORMÁTICA

Sistemas de Informação

Trabalho orientado pelo Prof. Doutor André Osório e Cruz de Azerêdo Falcão

2013



# Agradecimentos

Esta tese não representa apenas um projecto... Mas o projecto que foi o meu percurso académico. Por isso não me representa só a mim, representa também todos aqueles que contribuíram de algum modo para esta fase da minha vida. Desde já um obrigado.

Em primeiro lugar um agradecimento especial ao meu orientador Professor André Falcão pela oportunidade, apoio e confiança. Que devido ao entusiasmo que demonstra pelo mundo da química, acabou por fazer com que “simpatizasse” com o tema.

Quero também agradecer á minha mãe Alice, á minha tia Sofia, aos meus irmãos Fábio e Diogo e aos meus padrinhos Luís e Ília que estiveram sempre presentes, sempre prontos a apoiar, e que, mesmo nos momentos mais complicados conseguiram arrancar-me um sorriso.

Um agradecimento especial á Lili por todo o apoio, paciência, disponibilidade, preocupação e afecto demonstrado em todos os momentos.

Quero agradecer também aos meus amigos de Forles, sempre disponíveis para apoiar: Pedro, Cajó, Tó, Carlitos, Vitor, Bruno, Jurema, Luís, Miguel, Tiago Correia, Ti-TeX, Batatinha e Cenourinhas.

Quero agradecer também aos amigos: Esquina, Inês, Jorge, Paulo, Anderson, Beto, Daniel, Aida Pereira e Luís Filipe.

Quero também agradecer ao pessoal que conheci na faculdade, mas que são amigos pra vida: Insta, Epicurean, Paulino, Mac, Mano, Chimy, Tio Dark, Gonden, Dias, Krypton, Reis, Top-Ten, Manu, Insuportavel Milena, Pimpom, Kleo, Mascas, Gelásio, Copinho, Craveiro, Lopes e Zé.

Quero agradecer também á aquisição mais recente da família, o Matias.

**Obrigado a TODOS!**



*Dedico esta tese à minha mãe, à minha tia, aos meus irmãos, à Lili e ao Matias.*

## Resumo

A Barreira Hemato-Encefálica (BHE) é uma membrana que protege todo o sistema central de uma série de moléculas potencialmente nocivas para o funcionamento do cérebro. Apesar de ser um dos factores mais importantes a ter em conta para o desenvolvimento de novos fármacos, a investigação na área da permeabilidade da BHE é muito morosa e dispendiosa devido à dificuldade de testes *in vivo* e *in vitro*, os quais ainda impõem desafios técnicos consideráveis. São necessários até 16 anos para que um medicamento que actue no Sistema Nervoso Central chegue aos mercados e seja comercializado, sendo a sua permeabilidade na BHE um dos factores fundamentais para o seu sucesso.

Este trabalho procurou dar uma contribuição para o estudo da BHE sob duas vertentes principais. Em primeiro lugar procedeu-se à estruturação de uma aplicação web que permitisse a implementação e uso generalizado de um modelo de aprendizagem automática que permite estimar se um dado composto ultrapassa a BHE (*B<sub>3</sub>PP*). Em segundo lugar procedeu-se à implementação de um sistema de informação para entidades químicas para as quais existe na literatura informação documentada sobre a sua permeabilidade (*B<sub>3</sub>Info*). Estas aplicações foram estruturadas para permitir a que profissionais e investigadores nas áreas das ciências da vida, da saúde e da química possam utilizá-las sem dificuldade, entrando em linha de conta com as especificidades das representações moleculares, que se revestem de complexidades técnicas no seu uso, armazenamento e manipulação. Actualmente a base de dados do B3Info é populada por mais de 2000 compostos, com toda a informação química estrutural relevante incluindo as referências bibliográficas adequadas, bem como ligações directas a vários repositórios online. Inclui ainda mecanismos de pesquisa de compostos por semelhança estrutural e por subestrutura. A aplicação *B<sub>3</sub>PP* foi extensivamente testado pela comunidade científica internacional, tendo sido usado e mantido desde a sua publicação.

**Palavras-chave:** Barreira Hemato-Encefálica, B3Info, B3PP, compostos e permeabilidade.





# Abstract

The blood brain barrier (BBB) is a biological membrane that protects the Central Nervous System (CNS) against molecules potentially disruptive to the functioning of the brain. Despite the fact that it is one of the fundamental factors to account for when developing new drugs, research on the permeability of the BBB is challenging due to the difficulty of in vitro and in vivo testing, which still impose considerable technical difficulties and are very expensive. It is necessary about 16 years for the development of a drug that acts upon the CNS to reach the market, and the permeability of the BBB a critical factor for its success.

Thus work aims to help researchers in the study of the BBB in two main fronts. In the first place a web application was developed that puts to use a machine learning model that allows the estimation whether a given molecule is able to cross the BBB (B3PP). Secondly a web-based information system was structured for molecules for which there is literature data about its BBB permeability (B3Info). Both applications were developed to allow that professionals and researchers in the life and health sciences as well as chemists might be able to use them easily, by capturing the structural representations of molecules and specific query requirements that are totally different from typical information systems. Currently B3Info contains over 2000 molecules for which there is BBB verified information with all structural information including, bibliographic references for the relevant literature, as well as direct links to other online repositories. The system allows queries by similarity and substructure search. The B3PP application was extensively tested by the international community, having been actively maintained since it went public research in the Blood Brain Barrier (BBB).

**Keywords:** BBB, B3PP, B3Info, CNS and Blood Brain Barrier



# Conteúdo

Capítulo 1	Introdução.....	1
1.1	Motivação .....	1
1.2	Objectivos.....	3
1.3	Metodologia.....	3
1.4	Planeamento .....	4
1.5	Estrutura do documento.....	5
Capítulo 2	Trabalho relacionado.....	7
2.1	Grandes Sistemas de Informação .....	7
2.1.1	PubChem .....	7
2.1.2	Chemical Entities of Biological Interest (ChEBI).....	7
2.1.3	DrugBank .....	8
2.1.4	ChemSpider.....	8
2.2	ThermInfo.....	8
2.3	Resumo .....	8
Capítulo 3	Conceitos e tecnologias.....	11
3.1	Representação de informação química .....	11
3.2	Conceitos Químicos.....	12
3.3	Tecnologia Utilizadas .....	13
3.3.1	Desenvolvimento Web .....	13
3.3.2	Tratamento e Armazenamento de Dados .....	13
3.3.3	JChemPaint.....	13
3.3.4	CACTUS Web Service.....	13
3.3.5	Processamento de Informação Molecular .....	14
3.3.6	Linguagem R <sup>[4]</sup> .....	14
3.4	Arquitectura Web .....	14
Capítulo 4	Análise de requisitos .....	17
4.1	Público-alvo.....	17

4.2	<i>B<sub>3</sub>PP</i> .....	17
4.2.1	Análise do problema.....	17
4.2.2	Análise de requisitos .....	18
4.2.3	Casos de uso .....	18
4.2.4	Diagrama de actividades .....	19
4.3	<i>B<sub>3</sub>Info</i> .....	20
4.3.1	Análise do problema.....	20
4.3.2	Análise de requisitos .....	20
4.3.3	Casos de uso <i>B<sub>3</sub>Info</i> .....	21
4.3.4	Diagrama de actividades <i>B<sub>3</sub>Info</i> .....	22
4.4	Comunicação entre <i>B<sub>3</sub>Info</i> e <i>B<sub>3</sub>PP</i> .....	24
Capítulo 5	Desenvolvimento do <i>B<sub>3</sub>PP</i> .....	26
5.1	Arquitectura da Aplicação .....	26
5.2	Interface .....	27
5.3	Aplicação Web .....	28
5.3.1	Comunicação entre camadas .....	28
5.3.2	Validação de moléculas.....	28
5.3.3	Modelo de Predição.....	28
5.3.4	Informações químicas adicionais sobre a molécula introduzida.....	29
5.4	Predição da probabilidade de penetração de uma molécula na Barreira Hemato-Encefálica .....	29
Capítulo 6	Desenvolvimento do <i>B<sub>3</sub>Info</i> .....	32
6.1	Arquitectura da Aplicação .....	32
6.2	Base de Dados .....	33
6.3	Comunicação entre camadas .....	35
6.4	Frontend <i>B<sub>3</sub>Info</i> .....	35
6.4.1	Interface Frontend B3Info.....	35
6.4.2	Pesquisa Rápida.....	36
6.4.3	Pesquisa Estrutural .....	37

6.4.4	Pesquisa Subestrutural.....	38
6.4.5	Pesquisa por Similaridade .....	41
6.4.6	Download da Base de Dados.....	43
6.4.7	Secção de Ajuda .....	45
6.4.8	Comunicação com o <i>B<sub>3</sub>PP</i> .....	47
6.4.9	Área de administração .....	48
6.4.10	Funcionalidades Comuns .....	48
6.5	Backend .....	50
6.5.1	Interface Backend B3Info .....	50
6.5.2	Autenticação.....	51
6.5.3	Gestão de Moléculas .....	52
6.5.4	Gestão de Referências .....	56
6.5.5	Gestão da Associação Composto-Referencia.....	58
6.5.6	Gestão de Utilizadores .....	61
6.5.7	Actualizar versão da base de dados.....	63
6.5.8	Actualizar ficheiros auxiliares.....	63
6.6	Transformação e carregamento de dados .....	64
Capítulo 7	Conclusão .....	65
Capítulo 8	Bibliografia.....	68



# Lista de Figuras

Figura 1 - Dados de compostos químicos armazenados em folhas de cálculo .....	3
Figura 2 - Mapa de Gantt .....	5
Figura 3 - Protótipo de alta-fidelidade da interface do B3PP .....	27
Figura 4 - Arquitectura sistema de informação B3Info .....	32
Figura 5 - Diagrama de Classes UML do sistema B3Info .....	34
Figura 6 - Interface Frontend do sistema B3Info .....	35
Figura 7 - Pesquisa simples no sistema B3Info .....	36
Figura 8 - Exemplo de pesquisa simples pelo composto aspirin“ no <i>B<sub>3</sub>Info</i> .....	37
Figura 9 - Exemplo de pesquisa estrutural.....	37
Figura 10 - Exemplo de retorno de pesquisa estrutural .....	38
Figura 11 - Exemplo de uma pesquisa por SMARTS.....	38
Figura 12 - Resultado de uma pesquisa por subestrutura.....	39
Figura 13 - Ficheiro XML resultante de uma pesquisa por subestrutura.....	40
Figura 14 – Exemplo da funcionalidade do botão <i>More Info</i> .....	40
Figura 15 – Pesquisa por similaridade introduzindo um SMILES .....	41
Figura 16 – Pesquisa por similaridade através da applet java.....	41
Figura 17 – Pesquisa por similaridade através de SMILES.....	42
Figura 18 – Menu de escolha do tipo de ficheiro para download.....	43
Figura 19 – Ficheiro XML que contém toda a informação da base de dados .....	43
Figura 20 – Ficheiro SDF de download da base de dados .....	44
Figura 21 – Esqueleto do ficheiro XML de download da base de dados.....	45
Figura 22 – B3Info Help .....	45
Figura 23 – Funcionalidade de contacto com os administradores .....	46
Figura 24 – Constituição da equipa de desenvolvimento do B <sub>3</sub> Info.....	46
Figura 25 – Menu superior do sistema B3Info .....	47
Figura 26 – Resultado da previsão do composto <i>Aspirin</i> no B3PP .....	47
Figura 27 – Hiperligação para acesso ao backend do sistema B3Info.....	48
Figura 28 – Zoom sobre a imagem do composto.....	48



Figura 29 – Ligações externas associadas ao composto <i>Aspirin</i> .....	49
Figura 31 – Funcionalidade de inserir comentários sobre a molécula pesquisada .	49
Figura 32 – Menu de login do backend do sistema B <sub>3</sub> Info .....	51
Figura 33 . Funcionalidade de recuperar dados de acesso ao backend .....	51
Figura 34 – Funcionalidade de inserir molécula na base de dados .....	52
Figura 35 – Funcionalidade de <i>autocomplete</i> .....	53
Figura 36 – Funcionalidade de editar um composto .....	53
Figura 37 – Eliminar composto.....	54
Figura 38 – Inserção em massa .....	55
Figura 39 – Valida sequencia.....	55
Figura 40 – dicionar referência bibliográfica.....	56
Figura 41 - Editar uma referencia bibliográfica.....	57
Figura 42 – Eliminar referência .....	57
Figura 43 – Consulta lista de referências .....	58
Figura 44 – Adição de uma referência a um composto.....	58
Figura 45 – Edição da associação molécula-referência .....	60
Figura 46 – Eliminar associação Referencia-Molécula .....	60
Figura 47 – Criação de novo administrador .....	61
Figura 48 - Administrador já existente no sistema .....	61
Figura 49 – Remover administrador do sistema .....	62
Figura 50 – Edição do perfil .....	62
Figura 51 – Actualizar versão da base de dados .....	63
Figura 52 - Actualização de ficheiros .....	63
Figura 53 – Previsão da duração da actualização dos ficheiros .....	64



# Lista de Tabelas

Tabela 1 – Funcionalidades dos sistemas de informação de referência.....	9
Tabela 2 – Representação da informação química da <i>Aspirina</i> .....	12



# Lista de Siglas e Abreviaturas

<b>BHE</b>	Barreira Hemato-Encefálica
<b>LaSIGE</b>	Large-Scale Informatics Systems Laboratory
<b>B3PP</b>	Blood-Brain Barrier Penetration Prediction
<b>B3Info</b>	Blood-Brain Barrier Info
<b>IUPAC</b>	International of Pure and Applied Chemistry
<b>SMILES</b>	Simplified Molecular Input Line Entry System
<b>CASRN</b>	Chemical Abstracts Service Registry Number
<b>HTML</b>	Hypertext Markup Language
<b>MySQL</b>	My Structured Query Language
<b>PHP</b>	Hypertext Preprocessor

# Capítulo 1

## Introdução

No âmbito do segundo ano do Mestrado em Engenharia Informática, realizado no Departamento de Informática da Faculdade de Ciências da Universidade de Lisboa, mais concretamente no *LaSIGE* e com o objectivo de contribuir para a investigação da permeabilidade de um composto na Barreira Hemato-Encefálica foram desenvolvidos o sistema de informação *Blood-Brain Barrier Info (B3Info)* e a aplicação web *Blood-Brain Penetration Prediction (B3PP)*.

A aplicação web *B3PP* disponibiliza o acesso a um novo modelo que determina a probabilidade de uma molécula penetrar a BHE. Este acesso pode ser realizado através da *interface* web ou através do *web service*.

O *B3Info* é um sistema que disponibiliza informação estruturada e organizada sobre compostos químicos cuja penetração na BHE já foi testada e validada experimentalmente. Permite a realização de pesquisas de moléculas por nome comum, SMILES<sup>[6]</sup>, InChI, estrutura, subestrutura ou similaridade. Permite também o *download* da base de dados para uso pessoal.

### 1.1 Motivação

O processo para determinar se uma molécula penetra ou não a BHE por ensaios experimentais é muito dispendioso e moroso. <sup>[1]</sup> Perante isto surge a necessidade de prever a probabilidade de um composto penetrar ou não a BHE, e nesse sentido foi desenvolvido um novo modelo de previsão da permeabilidade na BHE. Para que esse modelo possa estar disponível a toda a comunidade científica era necessário o desenvolvimento de uma aplicação web que permitisse o acesso ao mesmo. Para tal foi desenvolvida a aplicação *Blood-Brain Barrier Penetration Prediction (B3PP)*.

A informação existente sobre os compostos cujos valores de penetração são conhecidos e foram criticamente avaliados encontra-se armazenada e dispersa em folhas

de cálculo (Figura 1). Este modo de armazenamento revela vários problemas, nomeadamente:

- **Integridade dos dados** – Para alterar ou apagar dados relativos a um composto em múltiplas linhas e/ou múltiplas folhas de calculo é necessário que a mesma acção seja repetida varias vezes. Com este procedimento, se algum campo é esquecido, os dados relativos ao composto vão tornar-se ambíguos e perder a integridade.
- **Redundância dos dados** – Este tipo de armazenamento é muitas vezes ineficiente, exigindo a repetição dos mesmos dados. Assim, o conjunto de dados cresce desnecessariamente, o acesso torna-se mais lento e a integridade dos dados fica mais difícil de controlar.
- **Vista limitada sobre os dados** – Não é possível a visualização de todo o conjunto de dados principalmente quando se encontram divididos por várias folhas de cálculo. Este aspecto é visível na Figura 1.1 onde apenas se consegue visualizar uma parte dos dados.
- **Capacidade de análise e performance** – À medida que o conjunto de dados vai crescendo, torna-se mais difícil de os analisar os dados e vão ser necessários mais recursos computacionais.
- **Partilha de dados** – A partilha dos documentos de cálculo pela comunidade torna-se muito difícil controlar pois qualquer alteração efectuada tem que ser novamente publicada por toda a comunidade.
- **Evolução** – A utilização de folhas de cálculo dificulta a evolução do conjunto de dados através de novas integrações.

A				D				E				A				B				C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
1	ordem	sigla	nome	p_nro	unidades																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													</

**Figura 1 - Dados de compostos químicos armazenados em folhas de cálculo**

## 1.2 Objetivos

O desenvolvimento da aplicação web  $B_3PP$  e do sistema de informação  $B_3Info$  tem como objectivo comum contribuir para o estudo da permeabilidade de um composto na Barreira Hemato-Encefálica.

Pretende-se com o desenvolvimento do  $B_3PP$  disponibilizar o acesso a um novo modelo de previsão da BHE. Para tal serão desenvolvidos uma *interface web* que permite ao utilizador a introdução de moléculas de modo a realizar a predição das mesmas e um *web service* para que o modelo possa ser acedido por aplicações externas.

O sistema de informação *B<sub>3</sub>Info* tem como objectivos principais organizar e disponibilizar a informação existente de forma estruturada, coerente e simples e permitir o *download* da base de dados do sistema para uso próprio. De modo a disponibilizar essa informação pretende-se desenvolver o *frontend* para permitir acesso à informação através de vários tipos de pesquisas e o *backend* para permitir a gestão de moléculas, referências e utilizadores.

### 1.3 Metodologia

Este projecto foi desenvolvido de acordo com as seguintes etapas:

- **Introdução** ao tema da permeabilidade na Barreira Hemato-Encefálica. Foi também formulado e analisado o problema e quais as tecnologias a utilizar.
- Etapa de **concepção** de soluções para a aplicação B3PP com base na análise de requisitos. Foram também desenvolvidos os protótipos de baixa fidelidade de modo a seleccionar a solução que mais apropriada.



- **Implementação da aplicação web B3PP** com base nos protótipos criados anteriormente. A interface da aplicação B3PP foi desenvolvida recorrendo às tecnologias HTML, CSS<sup>[8]</sup> e javascript. A camada de logica de negócio foi desenvolvida utilizando as tecnologias PHP<sup>[7]</sup> e python.
- **Avaliação** da aplicação web *B3PP* através de testes realizados por utilizadores. Após a fase de testes foram corrigidas algumas falhas detectadas durante esta fase.
- Etapa de **concepção** de soluções para o sistema de informação *B3Info* com base na análise de requisitos. Foram também desenvolvidos os protótipos de baixa fidelidade de modo a seleccionar a solução que mais apropriada.
- **Implementação do sistema de informação B3Info** de acordo com os protótipos criados anteriormente. Foram desenvolvidas as *interfaces* do *frontend* e do *backend* recorrendo às tecnologias HTML<sup>[16]</sup>, CSS e javascript. Foi desenvolvida a camada da lógica de negócio utilizando as tecnologias PHP e python. Foi também desenvolvida a camada de dados utilizando a tecnologia MySQL.
- Lançamento da versão *Alpha* e **Avaliação** do sistema B3Info através de testes realizados com utilizadores. Após a fase de testes foram corrigidas algumas falhas detectadas durante esta fase.
- **Lançamento da versão final** do B3Info.

## 1.4 Planeamento

Neste capítulo são apresentadas as diversas fases necessárias ao desenvolvimento desta tese. Tais fases encontram-se descritas no mapa de Gantt (Figura 2). A realização desta tese não aconteceu de acordo com o plano inicial de trabalho. Isso deveu-se a atrasos nas fases de testes e optimização de ambas as aplicações. Devido a problemas não ligados directamente à tese, houve um grande atraso no arranque da mesma, ou seja, apenas se iniciou em Dezembro de 2011 e também houve uma grande demora na escrita do relatório final.

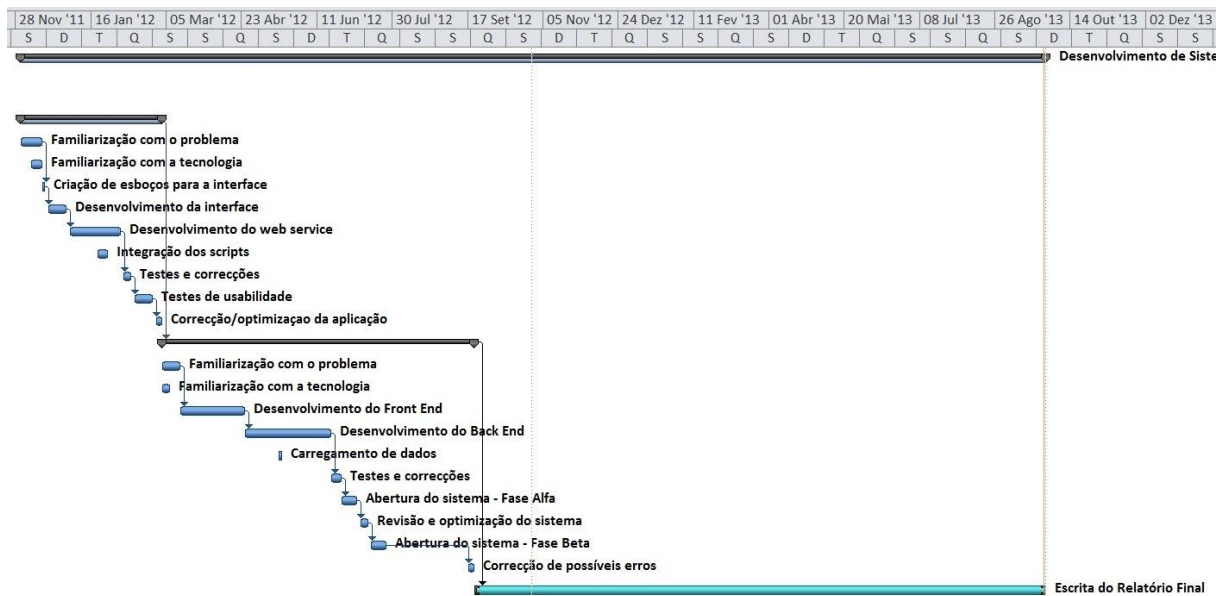


Figura 2 - Mapa de Gantt

## 1.5 Estrutura do documento

Este documento está organizado da seguinte forma:

- Capítulo 1 – Capítulo introdutório no qual se apresenta o contexto do trabalho, as motivações que levaram a realiza-lo, os contributos deste para o tema em questão e um breve resumo da estrutura do documento.
- Capítulo 2 – Capítulo informativo no qual se explicam o trabalho relacionados com o âmbito desta tese, grandes sistemas de informação e o ThermInfo.
- Capítulo 3 – Capítulo informativo no qual se explicam as tecnologias utilizadas no âmbito desta tese e a arquitectura web das aplicações.
- Capítulo 4 – Capítulo que descreve a análise de requisitos, casos de uso e diagramas de actividades do *B<sub>3</sub>PP* e do *B<sub>3</sub>Info*.
- Capítulo 5 – Capítulo que descreve todo o trabalho realizada para o desenvolvimento da aplicação web *B<sub>3</sub>PP*.
- Capítulo 6 – Capítulo que descreve todo o trabalho realizada para o desenvolvimento da aplicação web *B<sub>3</sub>Info*.
- Capítulo 7 – Capítulo de conclusão relativamente ao trabalho realizado, as dificuldades encontradas e trabalho futuro.



## Capítulo 2

### Trabalho relacionado

Neste capítulo serão abordados alguns dos grandes sistemas de informação com relevância no tema. Será também abordado o ThermInfo.

#### 2.1 Grandes Sistemas de Informação

##### 2.1.1 PubChem<sup>[13]</sup>

Lançado em 2004 pelo *National Center for Biotechnology Information* disponibiliza sinónimos, SMILES, InChI, propriedades e estruturas químicas sobre pequenas moléculas. O sistema PubChem agrega três bases de dados (*PubChem Substance*, *PubChem Compound* e *PubChem BioAssay*) e disponibiliza ferramentas de pesquisa por nome, pesquisa estrutural, pesquisa por similaridade e por subestrutura. Possibilita também o *download* da base de dados.

##### 2.1.2 Chemical Entities of Biological Interest (ChEBI)<sup>[12]</sup>

Consiste numa base de dados e ontologia de entidades moleculares focada em compostos pequenos. As entidades moleculares em questão são também produtos naturais ou sintéticos usados para intervir nos processos de organismos vivos. Permite pesquisas por nome comum, estrutura, similaridade e subestrutura. Disponibiliza também o download da base de dados.

### 2.1.3 DrugBank<sup>[11]</sup>

A base de dados DrugBank, disponibilizado pela Universidade de Alberta, é um recurso único de bioinformática e quimioinformática que se destaca pela enorme quantidade de informação farmacológica detalhada a respeito de compostos químicos. Permite pesquisas por nome comum, similaridade e subestrutura. Disponibiliza também o download da base de dados.

### 2.1.4 ChemSpider<sup>[10]</sup>

Lançado em 2007 pela Royal Society of Chemistry, é um sistema de informação sobre compostos químicos. Permite acesso a mais de 28 milhões de estruturas, propriedades e informações associadas. Integra e liga compostos de mais de 400 fontes de dados. Permite pesquisas por nome comum, estrutural, similaridade e subestrutura.

## 2.2 ThermInfo<sup>[9]</sup>

É um Sistema de Informação projectado e construído com o objectivo de recolher e recuperar valores termoquímicos criticamente avaliados. Na sua versão actual este permite recuperar o valor de uma propriedade termoquímica, introduzindo por exemplo, a estrutura molecular ou o nome de um composto. A estrutura deste sistema serviu de base para a concepção do *B3Info* e deste modo os seus métodos de pesquisa estrutural, pesquisa por subestrutura e por similaridade foram importados para o mesmo.

## 2.3 Resumo

Após a análise aos diferentes sistemas de informação mais relevantes no âmbito de informação sobre compostos químicos, identificaram-se as seguintes funcionalidades comuns a quase todos os sistemas: pesquisa por nome, pesquisa estrutural, pesquisa por subestrutura, pesquisa por similaridade e download da base de dados (Tabela 1).

No sistema de informação *B3Info* pretende-se implementar as funcionalidades mencionadas anteriormente e que são essenciais num sistema de informação de compostos químicos. Nesse sentido, pretende-se tais métodos em prol de uma base de dados desenvolvida para o efeito, ou seja, para armazenar valores de penetração de uma molécula na BHE, bem como as referências que os validam e outros dados estruturais e químicos associados á mesma.

	PubChem	ChEBI	DrugBank	ChemSpider	ThermInfo
Pesquisa Por nome	Sim	Sim	Sim	Sim	Sim
Pesquisa Estrutural	Sim	Sim	Sim	Sim	Sim
Pesquisa por subestrutura	Sim	Sim	Sim	Sim	Sim
Pesquisa por similaridade	Sim	Sim	Sim	Sim	Sim
Download da base de dados	Sim	Sim	Sim	Não	Não

**Tabela 1 – Funcionalidades dos sistemas de informação de referência**



## Capítulo 3

### Conceitos e tecnologias

De modo a facilitar a compreensão de alguns termos químicos utilizados neste documento, este capítulo faz uma breve descrição dos mesmos. Serão descritas também as tecnologias utilizadas.

#### 3.1 Representação de informação química

O **nome comum**<sup>[17]</sup>, que é atribuído a cada composto baseado numa nomenclatura sistemática de acordo com as recomendações da *IUPAC* é a forma mais fácil para o humano identificar uma molécula. Porém, para que uma máquina identifique correctamente uma molécula é necessário converter a estrutura química da mesma numa sequência de caracteres ASCII, ou seja, num **SMILES**. Esta especificação é legível para e as máquinas e para humanos, porém não garante unicidade, ou seja, duas moléculas podem gerar o mesmo SMILES. De modo a corrigir esta falha surgiu o **InChI**, que consiste numa cadeia de caracteres que deriva da representação estrutural da molécula. Apesar de o InChI garantir unicidade e também ser humanamente legível, a sua representação dificulta a pesquisa na web e em base de dados, e deste modo foi desenvolvido o **InChIKey** que consiste numa cadeia fixa de 25 caracteres resultante de aplicar o algoritmo SHA-256 à sequência InChI. A **fórmula molecular** indica os elementos presentes na molécula e o número de átomos de cada um desses elementos, o **peso molecular** representa a soma da massa atômica desses mesmos átomos e o **CASRN** é o identificador único instituído e atribuído a cada substância química pelo *Chemical Abstracts Service*. Este identificador não tem qualquer significado químico e é atribuído numa ordem sequencial, de forma a assegurar a unicidade. Também é possível representar graficamente uma molécula (Figura 3).

Na tabela 2 pode-se verificar todas as informações químicas mencionadas anteriormente relacionadas com a molécula *Aspirina*.



Aspirina	
Nome IUPAC	2-acetoxybenzoic acid
SMILES	<chem>O=C(Oc1ccccc1C(=O)O)C</chem>
InChI	InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)
InChIKey	BSYNRYMUTXBXSQ-UHFFFAOYSA-N
Fórmula Molecular	C <sub>9</sub> H <sub>8</sub> O <sub>4</sub>
Peso Molecular	180.157 g/mol
CASRN	50-78-2

Tabela 2 – Representação da informação química da *Aspirina*

## 3.2 Conceitos Químicos

Comparar compostos químicos não é uma tarefa fácil, porém a comparação é possível quando se descodifica o composto numa sequência de dígitos binários (bits). Esta sequência é considerada a impressão digital da molécula, ou seja, é a sua **fingerprint**. As **fingerprints** permitem comparar directamente moléculas e compreender o nível de similaridade entre elas, e para tal é utilizada a biblioteca *Open Babel*.

Quando temos um conjunto de compostos, poder realizar pesquisas que agrupem os mesmos por características comuns é uma funcionalidade essencial. Para tal é utilizada a tecnologia **SMARTS** pois permite realizar pesquisas subestruturais, ou seja, é possível especificar qual a subestrutura (átomos, número atómico, ligações, entre outras) comum a todas as moléculas retornadas na pesquisa. Esta tecnologia utiliza a notação proveniente dos SMILES, ou seja, átomos e ligações. Pode também utilizar a *wildcard* “\*” que representa qualquer átomo. Por exemplo, se for realizada a pesquisa pelo SMARTS “[#6]~[#6]” retornara todas as moléculas em que haja dois átomos ligados entre si, independentemente de qual o tipo de ligação.

O facto de uma molécula penetrar ou não a Barreira Hemato-Encefálica é medido pelo logaritmo do coeficiente de partição sangue-cérebro, também denominado de **LogBB**. Se  $\text{LogBB} < -1$ , o composto é considerado  $\text{LogBB}^-$ , ou seja, não passa a BHE. Se  $\text{LogBB} \geq -1$ , o composto é considerado  $\text{LogBB}^+$ , ou seja, passa a BHE.

### 3.3 Tecnologia Utilizadas

Este capítulo abrange as tecnologias utilizadas para desenvolvimento da aplicação *B3PP* e do sistema *B3Info*.

#### 3.3.1 Desenvolvimento Web

As interfaces do sistema de informação *B3Info* e da aplicação web *B3PP* foram desenvolvidas recorrendo às tecnologias **HTML** e **CSS** para definição da estrutura, cores e design das mesmas, á tecnologia **javascript** para manipulação da estrutura, adição e remoção de elementos para o utilizador obter feedback das suas acções e para inserir os resultados retornados pelo servidor, á tecnologia **Ajax** de modo a enviar e receber dados de forma assíncrona sem interferir com a exibição e comportamento da interface.

#### 3.3.2 Tratamento e Armazenamento de Dados

Toda a parte de recepção, tratamento e validação de dados encontra-se na camada de negócio. Esta foi desenvolvida utilizando a tecnologia **PHP** devido á facilidade de comunicação com a base de dados, e facilidade de comunicação integração com a tecnologia javascript que foi utilizada para desenvolvimento das interfaces. Para armazenar todos os dados moleculares e referências é necessário um SGBD. Para tal foi utilizado o **MySQL** <sup>[5]</sup> devido ao seu elevado desempenho, elevada fiabilidade e usabilidade e facilidade de integração com diversas linguagens de programação.

#### 3.3.3 JChemPaint

Além da representação por SMILES e/ou InChI também é possível representar graficamente uma molécula. Uma das ferramentas que inclui essa funcionalidade é a aplicação **JChemPaint**<sup>[15]</sup>, que disponibiliza uma applet java que permite ser integrada numa página web, ou seja, necessitando apenas do browser é possível construir graficamente uma molécula que a aplicação JChemPaint realiza a conversão da mesma para o SMILES correspondente e assim ser utilizada para o objectivo pretendido.

#### 3.3.4 CACTUS Web Service

Ao inserir um composto pelo nome é necessário realizar a tradução do nome no seu SMILES para que possa ser executado pelo modelo de permeabilidade na BHE, e para

tal utilizamos o web service do CACTUS<sup>[14]</sup>. Este serviço também é utilizado para obter a representação gráfica e sinónimos associados á molécula introduzida.

### 3.3.5 Processamento de Informação Molecular

Tecnologias de script como o **Python** são ideais para funcionalidades de análise de dados químicos. Porém, por questões de eficiência certas ferramentas de análise química são apenas desenvolvidos para outras tecnologias, como por exemplo o C++. Uma dessas ferramentas é o **Open Babel**, que possui extensas capacidades para ler e escrever formatos de arquivo molecular, manipulação de dados moleculares, pesquisa por SMARTS, suporta fingerprints, entre outras funcionalidades. Devido ao facto de o servidor que disponibiliza as aplicações *B3PP* e *B3Info* estar preparado para execução da tecnologia Python e de necessitarmos das funcionalidades disponibilizadas pela ferramenta Open Babel, foi utilizado o **Pybel**<sup>[3]</sup>, que consiste num módulo Python que permite o acesso às funcionalidades químicas disponibilizadas pelo Open Babel. Assim podemos desenvolver um script em Python para manipulação de dados químicos.

### 3.3.6 Linguagem R<sup>[4]</sup>

A tecnologia *R* é uma linguagem e ambiente de desenvolvimento que permite correr modelos. Tecnologia é principalmente utilizada para manipular informação, cálculos e manipulação gráfica. Esta foi utilizada para desenvolver o modelo de predição disponibilizado pelo *B3PP*.

## 3.4 Arquitectura Web

Tendo como objectivo o desenvolvimento de duas aplicações web de simples e directo acesso, onde as únicas ferramentas que o utilizador necessitaria seriam um *browser* e acesso á internet, o modelo escolhido para as aplicações *B3PP* e *B3Info* foi o modelo cliente-servidor (Figura 3). É um modelo onde o processamento da informação é dividido em módulos ou processos distintos. Um processo é responsável pela manutenção da informação (Servidor), enquanto o outro é responsável pela obtenção dos dados (Cliente).

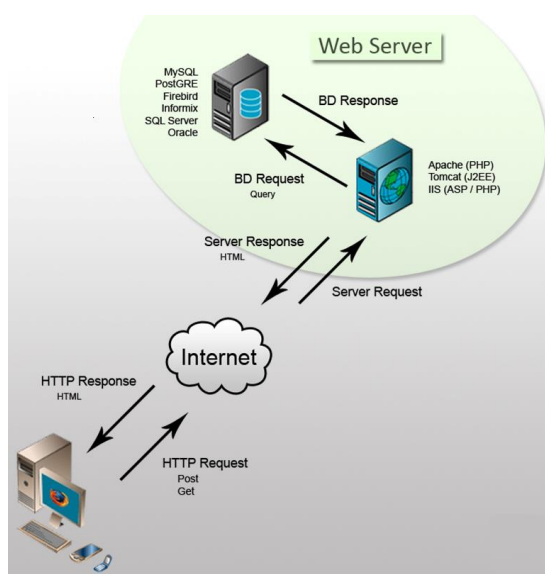
Esta arquitectura consiste em 3 componentes principais:

- **Servidor:** Equipamento que disponibiliza serviços para os clientes, ou seja, é a infra-estrutura onde estão alojadas as aplicações *B3PP* e *B3Info* e que disponibiliza todas as funcionalidades das mesmas. No caso do *B3Info* o

servidor inclui também a base de dados de compostos químicos necessária para o funcionamento do mesmo.

- **Cliente:** Computador que está ligado á rede de modo a aceder aos serviços fornecidos pelo servidor, neste caso serão todos os utilizadores que pretendam aceder às funcionalidades das aplicações, necessitando apenas de um *browser* para o efeito.
- **Rede:** Meio de transporte por onde passam os dados que vão do servidor para o cliente e vice-versa, neste caso será a internet.

Neste projecto foi utilizada a arquitectura LAMP (Linux, Apache HTTP Server, MySQL, PHP ou Python) que é uma arquitectura baseada no modelo cliente-servidor.



**Figura 3 – Architectuta web**

([http://flavioaf.files.wordpress.com/2010/02/arquitetura\\_web.jpg](http://flavioaf.files.wordpress.com/2010/02/arquitetura_web.jpg))



## Capítulo 4

### Análise de requisitos

As principais ilações tiradas aquando da recolha do trabalho relacionado foram consideradas nesta fase, tanto a nível dos sistemas de informação mencionados anteriormente como a nível de documentação existente. Foram também realizadas discussões com potenciais utilizadores dos sistemas *B3PP* e *B3Info*, de modo a clarificar o problema e discutir propostas de resolução.

Os **requisitos funcionais** definem as funções do *software*, isto é, os cálculos, detalhes técnicos, processamento e manipulação de dados e outras funcionalidades, ou seja, especificam o funcionamento do sistema. Os **requisitos não funcionais** abordam os critérios que julgam o sistema em termos de desempenho, usabilidade, confiabilidade, segurança, disponibilidade, manutenção e tecnologias envolvidas. Os requisitos não funcionais podem constituir restrições aos requisitos funcionais.

#### 4.1 Público-alvo

O público-alvo dos sistemas *B3PP* e *B3Info* são principalmente pessoas com formação em Química e que trabalham na área da investigação da permeabilidade de compostos na BHE. Assim, estamos perante utilizadores altamente qualificados na análise dos dados disponibilizados sobre compostos químicos e sobre a BHE.

#### 4.2 *B<sub>3</sub>PP*

##### 4.2.1 Análise do problema

Tal como foi descrito no capítulo 1, era necessário o desenvolvimento de uma aplicação web que permita o acesso a um novo modelo de previsão de penetração de um composto na BHE. Esse acesso deve ser possível através de uma *interface web*

disponibilizada para o utilizador e por um *web service* para o acesso por parte de aplicações.

### 4.2.2 Análise de requisitos

Após analisar devidamente o problema e com a participação de elementos que se enquadram no conceito de público-alvo foi possível identificar quais os requisitos associados à aplicação *B<sub>3</sub>PP*.

- **Requisitos funcionais:**

- ✓ Deve ser possível a predição de um composto através da inserção do nome, SMILES ou InChI
- ✓ Comunicação entre o cliente e servidor deve ser efectuada através de Web Service

- **Requisitos não funcionais:**

- ✓ Retorno do resultado da predição não deve ser demorado
- ✓ A informação proveniente do *Web Service* deve ser coerente com a apresentada pela interface
- ✓ A interface deve ser simples, intuitiva e focada no objectivo de modo a diminuir o tempo de aprendizagem necessário ao utilizador
- ✓ O resultado da predição deve ser apresentado de modo facilmente legível e coerente de forma a facilitar o seu entendimento
- ✓ Deverá ser compatível com todos os *browsers* de modo a não haver limitações de acesso
- ✓ Deverá permitir ao utilizador aceder à versão pretendida
- ✓ Deverá permitir a actualização/modificação do modelo de predição sem ser necessário alterar a estrutura/funcionalidade da aplicação

### 4.2.3 Casos de uso

A Figura 4 representa as funcionalidades identificadas para a aplicação web *B<sub>3</sub>PP* e os respectivos actores.

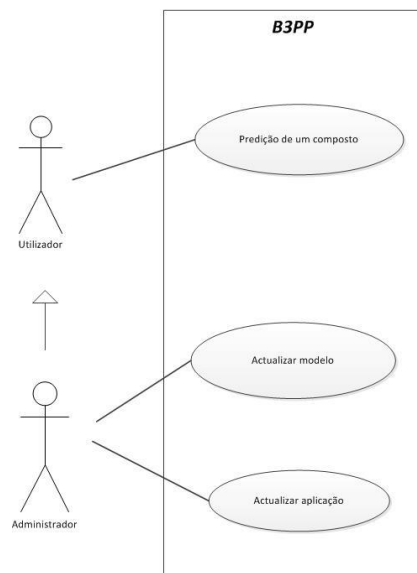


Figura 4 – Casos de uso  $B_3PP$

#### 4.2.4 Diagrama de actividades

O diagrama de actividades representado na figura5, representa a predição de uma molécula na aplicação B3PP. O utilizador insere uma molécula através da *interface web*, esta envia a molécula para a aplicação onde vai ser processada e posteriormente retornado o resultado.

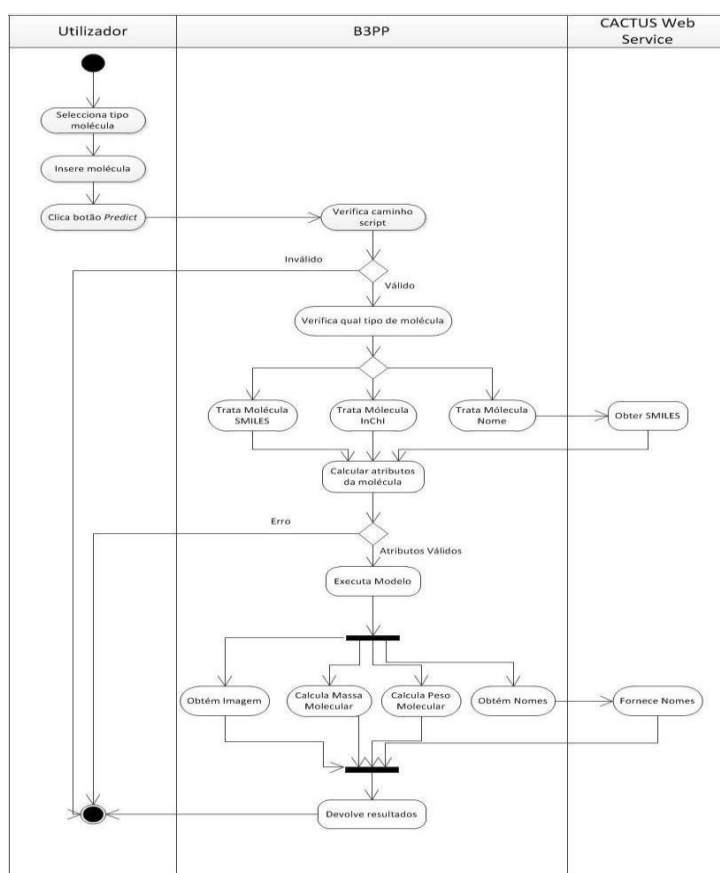


Figura 5 – Diagrama de actividades  $B_3PP$



## 4.3 *B<sub>3</sub>Info*

### 4.3.1 Análise do problema

É necessário o desenvolvimento de um Sistema de Informação que possibilite o acesso a informações de compostos químicos sobre os quais existem valores de penetração na BHE conhecidos e criticamente avaliados. Estes valores encontram-se armazenados em ficheiros de cálculo e é necessário a sua estruturação sob a forma de base de dados relacional. O sistema deverá permitir a gestão e manutenção dos dados associados ao mesmo, para tal serão necessários o *frontend* de modo a permitir aos utilizadores realizarem pesquisas e o *backend* para gestão de conteúdos por parte dos administradores do sistema.

### 4.3.2 Análise de requisitos

Após analisar devidamente o problema e com a participação de elementos que se enquadram no conceito de público-alvo foi possível identificar quais os requisitos associados à aplicação *B<sub>3</sub>PP*.

▪ **Requisitos funcionais:**

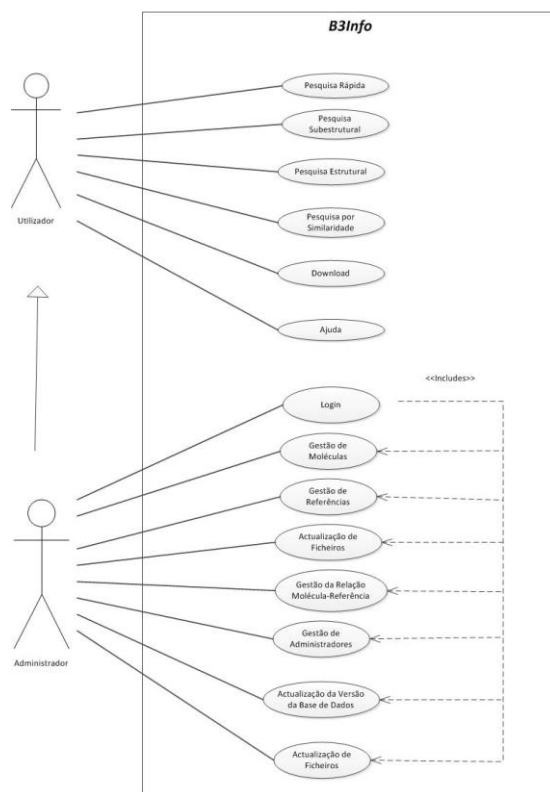
- ✓ Deverá permitir pesquisar compostos através do nome, SMILES, InChI ou fórmula molecular
- ✓ Deverá permitir pesquisar compostos pela estrutura e subestrutura molecular e similaridade
- ✓ Deve ser possível realizar o *download* da base de dados para ficheiro XML e SDF
- ✓ Possibilidade de reportar erros e opinião dos utilizadores
- ✓ Deverá ser possível inserir comentários relativos às moléculas
- ✓ Possibilitar a gestão das moléculas através da inserção, actualização e remoção das mesmas
- ✓ Possibilitar a gestão de referencias através da inserção, actualização e remoção das mesmas
- ✓ Possibilitar a gestão de utilizadores através da inserção, actualização e remoção dos mesmos
- ✓ Deverá ser possível actualizar a versão do sistema

▪ **Requisitos não funcionais:**

- ✓ Garantir integridade dos dados armazenados na base de dados
- ✓ A interface deve ser simples, intuitiva e focada no objectivo de modo a diminuir o tempo de aprendizagem necessário
- ✓ A adição ou modificação de funcionalidades não deve interferir com as restantes funcionalidades
- ✓ O retorno do resultado não deve ser demorado
- ✓ O resultado deve ser apresentado de forma legível e coerente de forma a facilitar o entendimento
- ✓ Os dados devem ser armazenados numa base de dados com arquitectura relacional adequada para que todas as operações se processem de forma rápida, eficiente e sem exigir elevados custos computacionais

### 4.3.3 Casos de uso *B<sub>3</sub>Info*

A Figura 6 representa as funcionalidades identificadas para o sistema de informação *B<sub>3</sub>Info* e os respectivos actores.



**Figura 6 – Casos de uso *B<sub>3</sub>Info***

#### 4.3.4 Diagrama de actividades *B<sub>3</sub>Info*

Devido á semelhança das acções necessárias para a realizar a pesquisa de uma molécula no sistema, as funcionalidades *Pesquisa Rápida* e *Pesquisa Estrutural* foram agrupadas no mesmo diagrama de actividade (Figura 7). O mesmo sucedeu com as funcionalidades *Pesquisa por Similaridade* e *Pesquisa Subestrutural* (Figura 8).

O utilizador insere o nome, SMILES, InChI, formula ou estrutura molecular através da interface e esta é pesquisada na base de dados associada ao B3Info. Se existir no sistema este retorna toda a informação associada á molécula introduzida, caso contrario, retorna um aviso a informar o utilizador que a molécula não existe no sistema.

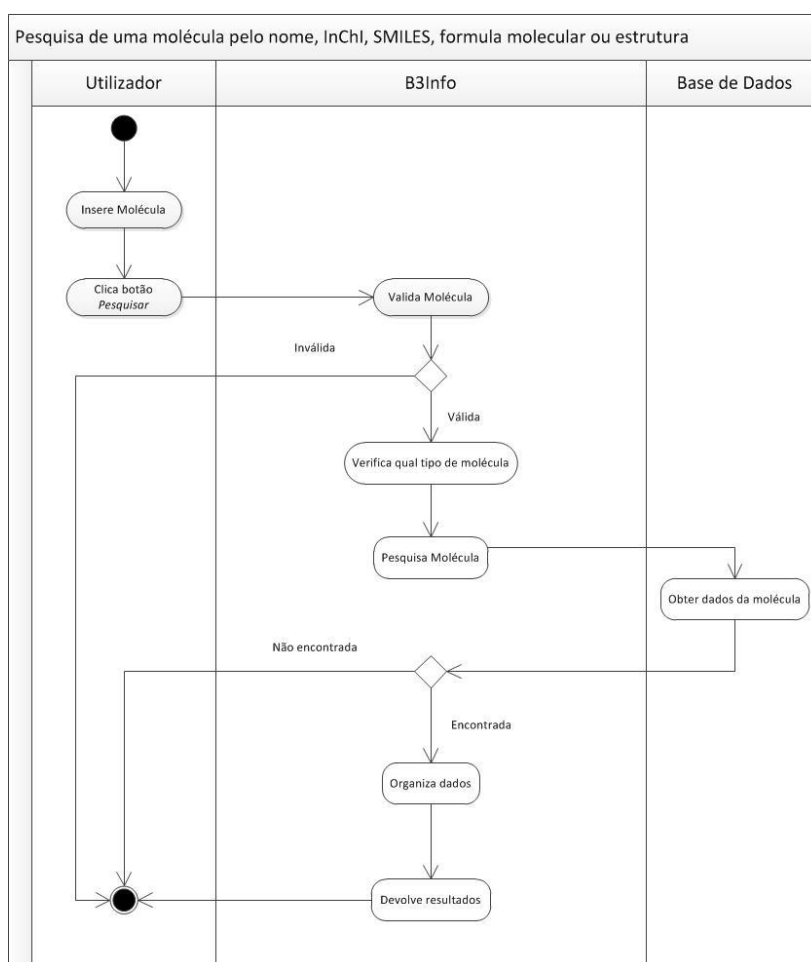
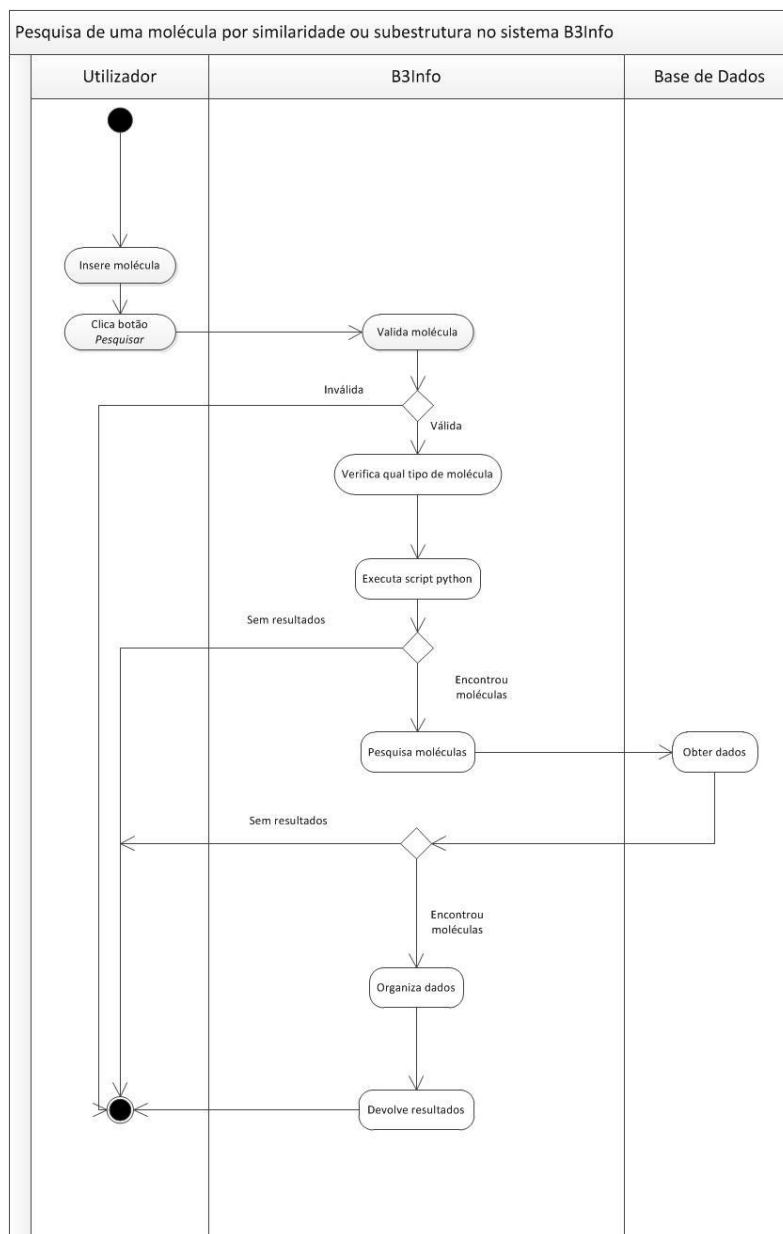


Figura 7 - Diagrama de actividade para pesquisa de uma molécula no B3Info

A Figura 8 representa as acções necessárias para pesquisa de uma molécula por similaridade ou por subestrutura. O utilizador insere a molécula, esta é validada e executada nos scripts desenvolvidos em python. De seguida são obtidos os dados relacionados com as moléculas devolvidas pelo script e estes dados são retornados ao utilizador.



**Figura 8 - Diagrama de actividade para pesquisa por similaridade ou subestrutura no B3Info**

A Figura 9 representa as acções necessárias para realizar a funcionalidade de download das moléculas da base de dados. O utilizador escolhe qual o formato do ficheiro que pretende e o sistema fornece o ficheiro contendo todos os dados contidos na base de dados.

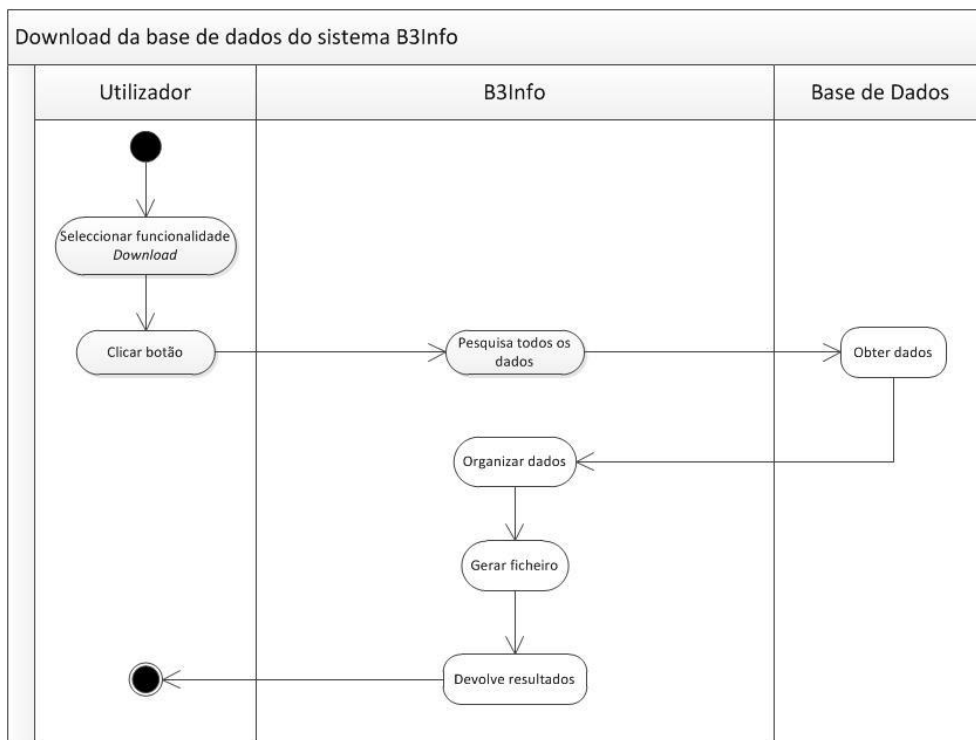


Figura 9 - Diagrama de actividade de download da base de dados do sistema B3Info

#### 4.4 Comunicação entre *B<sub>3</sub>Info* e *B<sub>3</sub>PP*

A comunicação entre o *B<sub>3</sub>Info* e o *B<sub>3</sub>PP* será possível através de 2 funcionalidades: haverá uma hiperligação no menu superior do *B<sub>3</sub>Info* que abrirá num novo separador a aplicação *B<sub>3</sub>PP* e em cada pesquisa será possível executar a molécula pesquisada no modelo através da invocação do web service disponibilizado pelo *B<sub>3</sub>PP*. Esta invocação é feita de modo assíncrono sem e depois o resultado integrado na vista sem que seja necessário carregar novamente a mesma.



## Capítulo 5

### Desenvolvimento do $B_3PP$

Como mencionado anteriormente, o objectivo desta aplicação é possibilitar o acesso a um novo modelo de previsão da permeabilidade de um composto na BHE. Neste capítulo serão descritos todos os processos que foram realizados para o desenvolvimento do  $B_3PP$ .

#### 5.1 Arquitectura da Aplicação

Uma arquitectura de *software* em camadas é mais simples de implementar e traz grandes benefícios, principalmente porque permite abstrair a lógica de negócio dos diferentes módulos do sistema. Esta organização facilita a manutenção, a portabilidade e a escalabilidade, factores importantes quando queremos partilhar funcionalidades e informação entre aplicações. Com base nestes factores, esta foi a arquitectura escolhida para o  $B_3PP$ .

Como demonstra a Figura 10, a aplicação  $B_3PP$  é composta por **camada de apresentação** (interface) que interage com o utilizador e pela **camada de lógica de negócio** que trata da validação dos compostos inseridos, da previsão do modelo e da recolha de informações associadas ao composto que irão enriquecer a interface.



Figura 10 - Arquitectura da aplicação  $B_3PP$

## 5.2 Interface

Facilidade de interacção aliada a uma boa apresentação são ingredientes fundamentais no sucesso de uma aplicação. De forma a oferecer ao utilizador as funcionalidades identificadas na etapa de *Análise de Requisitos*, deve haver algumas preocupações na implementação da *interface*. Este é um processo iterativo, na medida em que uma boa interface necessita de vários passos e ciclos para se adaptar às necessidades dos mesmos.

Estas tarefas foram realizadas através das seguintes formas:

- **Protótipos de baixa-fidelidade:** com o intuito de avaliar o desenho de forma simples e pouco dispendiosa em termos de tempo e trabalho, foram criados alguns protótipos de baixa-fidelidade até ser encontrado o protótipo que se enquadrava nos requisitos identificados anteriormente.
- **Protótipos de alta-fidelidade:** Depois de validado o protótipo de baixa-fidelidade, procedeu-se á construção do protótipo de alta-fidelidade (Figura 11) que representa o aspecto final da aplicação. O seu desenvolvimento focou-se principalmente no aspecto do design.

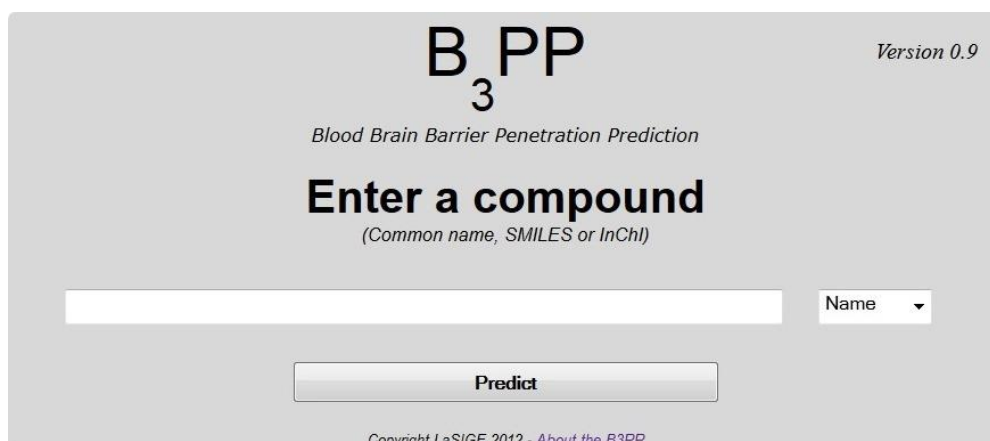


Figura 11 - Protótipo de alta-fidelidade da interface do B3PP

A *interface* é responsável numa primeira fase, pela recolha dos compostos químicos inseridos pelo utilizador e envio destes para a camada de lógica de negócio. Posteriormente também é responsável pela recepção da informação proveniente da camada de lógica de negócio e integração da mesma. Esta foi desenvolvida utilizando a tecnologia **CSS** para definir o estilo da *interface*, cores de fundo, estilo de letra, tamanha das imagens, entre outras, **HTML** para desenvolvimento do *design* da *interface* e **javascript** para manuseamento e alteração do aspecto da *interface*, mais propriamente por providenciar *feedback* ao adicionar o estado de *Processing* após clique



no botão *Predict* e posterior inserção na mesma da informação relativa á molécula retornada pelo servidor. A interface produzida vai de encontro aos requisitos identificados anteriormente, ou seja, é simples, intuitiva e focada no objectivo. É também compatível com os *browsers* mais utilizados.

## 5.3 Aplicação Web

Neste capítulo serão apresentados os detalhes relacionados com o desenvolvimento da aplicação web, ou seja, é definido como as moléculas são processadas, quais os *inputs* que são válidos, quais as informações a recolher sobre cada molécula e a execução do modelo de previsão da permeabilidade da BHE em relação á mesma.

### 5.3.1 Comunicação entre camadas

A comunicação entre a interface e a aplicação web é realizada de forma assíncrona através da tecnologia *javascript* do lado do cliente e serviços RESTful do lado do servidor. Deste modo é possível fornecer *feedback* ao utilizador após este fazer o pedido de predição, através da adição do estado de *processing* á interface e posteriormente a inserção da informação proveniente da camada de lógica de negócio relativa á molécula introduzida, de forma dinâmica. Para que esta comunicação assíncrona seja possível através do *javascript*, utilizamos o objecto *XMLHttpRequest*. Este envia um pedido HTTP para o servidor e aguarda uma resposta do mesmo. No caso do *B3PP* a resposta consiste num ficheiro XML com informações moleculares e resultado da execução do modelo a integrar na *interface*.

### 5.3.2 Validação de moléculas

Após as moléculas serem inseridas pelo utilizador, é necessário a validação das mesmas. Para tal, é utilizado um script desenvolvido em *python* que recebe como parâmetros o InChI ou o SMILES e efectua essa validação recorrendo às funcionalidades do Pybel. Se a molécula for válida, o script retorna atributos da mesma que irão ser consumidos pelo modelo de predição da permeabilidade, caso contrário retorna erro.

### 5.3.3 Modelo de Predição

Com o objectivo de contribuir para a diminuição do tempo necessário de investigação de novos compostos que podem ou não penetrar na Barreira Hemato-Encefálica, o grupo XLDB pertencente á unidade de investigação LaSige desenvolveu

um novo modelo para determinar a probabilidade de um composto atravessar a BHE. Este foi desenvolvido em *R* e tem como base estatística Bayesiana.

A evolução no aperfeiçoamento do modelo de predição ficará registada por versões. Estas ficarão armazenadas no servidor e poderão ser acedidas através da invocação directa do *web service*, ou seja, sem a utilização da interface. Actualmente encontra-se na versão 0.9.

### 5.3.4 Informações químicas adicionais sobre a molécula introduzida

De forma a enriquecer a experiência dos utilizadores ao utilizarem a aplicação *B3PP*, a interface é enriquecida com informações adicionais sobre a molécula introduzida. Estas são recolhidas na camada de lógica de negócio e enviadas para a *interface* no ficheiro XML enviado através do *web service*. A **fórmula e peso moleculares** obtêm-se através de um script desenvolvido em *python*, que recebe como parâmetro de entrada um InChI ou SMILES e que devolve a respectiva fórmula e peso moleculares, os **nomes comuns** associados á molécula introduzida obtêm-se através do *webservice* do CACTUS que tem como parâmetro de entrada o nome da molécula, o SMILES ou o InChI e retorna os nomes comuns associados á mesma e para a **representação gráfica da molécula** é preparado o URL que permite obter a imagem associada á molécula introduzida através do *webservice* do CACTUS. Este URL após recebido pela interface, é carregado directamente pela mesma de forma a apresentar a imagem correspondente á molécula introduzida.

## 5.4 Predição da probabilidade de penetração de uma molécula na Barreira Hemato-Encefálica

Como mencionado anteriormente, o objectivo maior desta aplicação é proporcionar o acesso a um novo modelo de previsão de permeabilidade na BHE. Tal acesso é proporcionado através de uma *interface web* e de um *webservice*.

A **interface web** encontra-se disponível através do URL *b3pp.lasige.di.fc.ul.pt* e ao aceder a esta, ao utilizador apenas necessita de inserir a molécula, escolher o identificador e pressionar o botão “*Predict*”. Após esta sequência a aplicação entra em processamento e como se pode verificar na Figura 12 retorna o resultado do modelo juntamente com mais alguma informação molecular associada á molécula inserida.

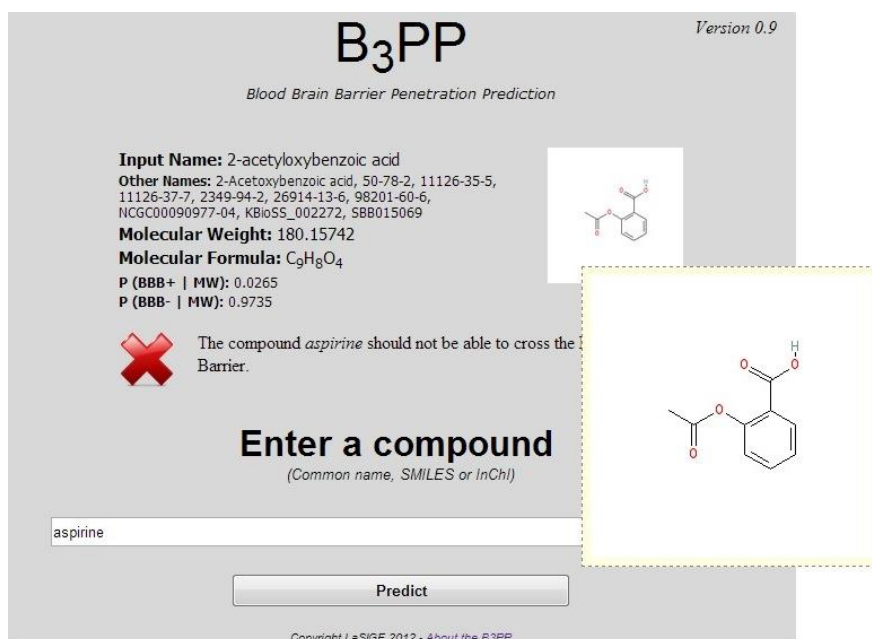


Figura 12 - Interface da aplicação B3PP após predição para a molécula *aspirine*

O *webservice* disponibilizado pela aplicação B3PP através do URL *b3pp.lasige.di.fc.ul.pt/webservices/mol.php/type/version/molecule* além de possibilitar o acesso ao modelo nas mesmas condições que a *interface web*, permite também escolher qual a versão do modelo que se pretende aceder. A Figura 13 mostra o conteúdo do ficheiro XML proveniente da aplicação.

```
<?xml version="1.0" encoding="UTF-8" ?>
<results>
  <result>
    <nome>2-acetyloxybenzoic acid</nome>
    <outro_nome>
      2-Acetoxybenzoic acid, 50-78-2, 11126-35-5, 11126-37-7, 2349-94-2, 26914-13-6, 98201-60-6, NCGC00090977-04, KBioSS_002272, SBB015069
    </outro_nome>
    <imagem>
      http://cactus.nci.nih.gov/chemical/structure/Aspirine/image
    </imagem>
    <mw>180.15742</mw>
    <formula>C9H8O4</formula>
    <P1>0.02649915</P1>
    <P2>0.9735008</P2>
    <valor>1</valor>
  </result>
</results>
```

Figura 13 - Conteúdo do ficheiro XML proveniente da aplicação



## Capítulo 6

### Desenvolvimento do *B<sub>3</sub>Info*

O *B<sub>3</sub>Info* é um Sistema de Informação que reúne informação validada experimentalmente sobre moléculas que penetram ou não a Barreira Hemato-Encefálica. Este torna-se único no estudo da penetração na BHE pois é um sistema de acesso público que contém medições reais sobre os valores da permeabilidade das moléculas na BHE.

#### 6.1 Arquitectura da Aplicação

A organização em camadas distintas (Figura 14) torna o sistema mais flexível permitindo que as partes possam ser modificadas de forma independente e sem prejuízo para as restantes. Facilita também a manutenção, portabilidade e escalabilidade do sistema.

De modo a que o sistema de informação *B<sub>3</sub>Info* tire partido das vantagens de uma arquitectura em camadas, este é constituído pela **camada de apresentação** que inclui todas as interfaces do sistema, pela **camada de lógica de negócio** que é responsável pelo tratamento e validação dos dados inseridos, organização dos dados de pesquisa a apresentar ao utilizador, acesso á base de dados, entre outras, e pela **camada de dados** que armazena todas as informações relativas a moléculas, referências, utilizadores e versões do sistema.

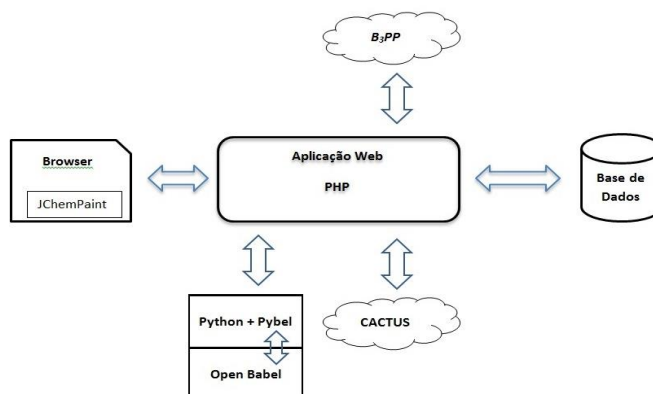


Figura 14 - Arquitectura sistema de informação B<sub>3</sub>Info

## 6.2 Base de Dados

A estrutura da base de dados foi concebida para permitir o acesso rápido, acomodar dados heterogéneos e manter a integridade referencial dos mesmos. O conjunto de dados é constituído por propriedades estruturais e químicas de moléculas, seleccionados de literatura científica relevante, assim como as respectivas referências bibliográficas. Foi desenvolvido um modelo relacional constituído por entidades que contêm os atributos que as caracterizam e pelas relações entre elas. Cada linha da entidade (tuplo) representa uma colecção de dados relacionados. O modelo relacional foi escolhido uma vez que tem um bom desempenho e facilidade em realizar consultas aos dados (traves da linguagem *MySQL*), é fácil de administrar, tem uma ampla aceitação e encontra-se bem documentado.

A Figura 15 mostra o diagrama de classes (UML) da estrutura da base de dados e as tabelas encontram-se definidas da seguinte forma:

- *compound*: tabela responsável pelo armazenamento dos compostos, contém 7 atributos (*c\_id*, *inchi*, *inchi\_key*, *canonical\_smiles*, *casrn*, *formula* e *mw*) dos quais o *inchi\_key* e o *c\_id* são únicos e este ultimo atribuído automaticamente pelo MySQL.
- *Reference*: tabela responsável pelo armazenamento de referências, que contém 12 atributos (*r\_id*, *authors*, *title*, *volume*, *issue*, *year*, *publisher*, *url*, *begin\_oage*, *end\_page*, *doi* e *ref\_all*) dos quais *r\_id* é único e atribuído automaticamente pelo MySQL.
- *compound\_data\_ref*: tabela respnsavel por armazenar as associações eferencia-molecula, contém 5 atributos (*bbb\_boolean*, *logbb*, *obervation*, *fk\_compound\_id* e *fk\_reference\_id*) dos quais a combinação *fk\_compound\_id*, *fk\_reference\_id* é única pois é chave primária, *fk\_compound\_id* é chave estrangeira para o atributo *c\_id* da entidade *compound* e *fk\_reference\_id* é chave estrangeira para o atributo *r\_id* da entidade *reference*.
- *compound\_name*: tabela responsável pelo armazenamento dos nomes dos compostos, contém 3 atributos (*n\_id*, *name* e *fk\_compound\_id*) dos quais *n\_id* é único e atribuído automaticamente pelo MySQL e *fk\_compound\_id* é chave estrangeira para o atributo *c\_id* da entidade *compound*.
- *compound\_image*: tabela responsável pelo armazenamento da correspondencia entre o ficheiro e o composto, contém 3 atributos (*img\_id*, *file\_name* e *fk\_compound\_id*) dos quais *img\_id* é único e atribuído automaticamente pelo MySql e *fk\_compound\_id* é chave estrangeira para o atributo *c\_id* da entidade *compound*.

- *compound\_link*: tabela responsável pelo armazenamento dos links externos dos compostos, contém 4 atributos (l\_id, entity, link e fk\_compound\_id) dos quais l\_id é único e atribuído automaticamente pelo MySQL e fk\_compound\_id é chave estrangeira para o atributo c\_id da entidade compound.
- *compound\_user\_comment*: tabela responsável pelo armazenamento dos comentários inseridos pelos utilizadores, contém 3 atributos (com\_id\_comment e fk\_compound\_id) dos quais com\_id é único e atribuído automaticamente pelo MySQL e fk\_compound\_id é chave estrangeira para o atributo c\_id da entidade compound.
- *db\_version*: tabela responsável pelo registo da versão da base de dados, contém 3 atributos (v\_id, version e user\_add) dos quais version e v\_id são únicos e este ultimo atribuído automaticamente pelo MySQL.
- *User*: tabela responsável pelo armazenamento dos registos dos utilizadores para poderem realizar login, contém 5 atributos (u\_id, username, password, name e email) dos quais username, email e u\_id são únicos e este ultimo atribuído automaticamente pelo MySQL.

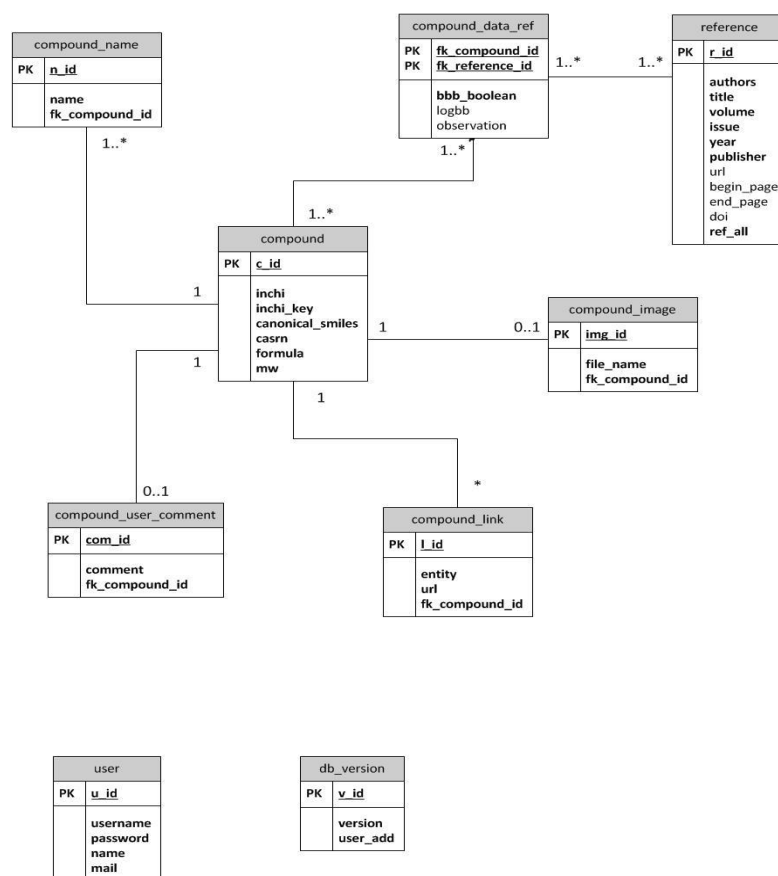


Figura 15 - Diagrama de Classes UML do sistema B3Info

## 6.3 Comunicação entre camadas

A comunicação entre as várias interfaces e a aplicação web é realizada através da tecnologia javascript que realiza pedidos assíncronos através do protocolo HTTP, ou seja, é indicado o URI do recurso, quais as variáveis a incluir do *POST* HTTP e a função que irá ser executada quando receber a resposta. Do lado do servidor, recebe essas variáveis no PHP e realiza o processamento previsto pela função. Após o processamento do pedido, o resultado do mesmo é retornado á função javascript definida a ser executada quando o servidor respondesse. A comunicação entre o sistema e a base de dados é realizada através de funções PHP disponibilizadas para o efeito, de início é realizada a conexão com a base de dados e depois é produzida a query SQL a executar sobre a BD. Após a execução a query na BD o resultado é processado no servidor através do PHP.

## 6.4 Frontend *B<sub>3</sub>Info*

É através do *frontend* que o utilizador pode interagir com o sistema B3Info e aceder às funcionalidades de pesquisa disponibilizadas. Este disponibiliza também a possibilidade de o utilizador realizar download da base de dados para uso pessoal, inserir comentários nas moléculas, secção de ajuda com explicação sobre as funcionalidades e contactos dos administradores.

### 6.4.1 Interface Frontend B3Info

Interface (Figura 16) desenvolvida a pensar no utilizador, com as funcionalidades acessíveis de forma rápida e intuitiva para que o tempo de aprendizagem necessário seja reduzido e que as informações relativas a compostos sejam apresentadas estruturadas e de forma coerente.

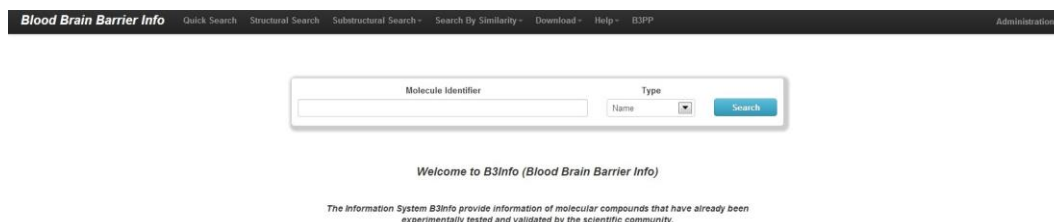


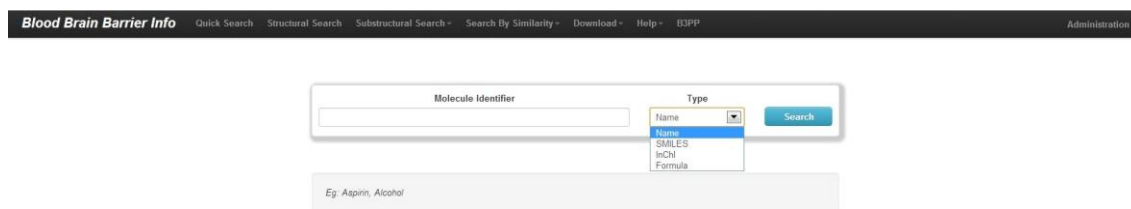
Figura 13 - Interface Frontend do sistema B3Info



Esta foi desenvolvida recorrendo às tecnologias HTML e CSS para definição e preenchimento do layout, menus, botões, cores, estilos de letra e animações. Foi também utilizada a tecnologia javascript para modificação dos elementos HTML, ou seja, adição de estados de *loading* para fornecer feedback ao utilizador e posterior introdução na página web dos resultados retornados pelo servidor.

### 6.4.2 Pesquisa Rápida

A funcionalidade de *Quick Search* (Figura 6.X) permite ao utilizador realizar uma pesquisa de uma molécula no sistema através do seu nome comum, SMILES, InChI ou fórmula molecular.



The screenshot shows the 'Blood Brain Barrier Info' website header with navigation links: Quick Search, Structural Search, Substructural Search, Search By Similarity, Download, Help, and B3PP. Below the header is a search form with a 'Molecule Identifier' input field, a 'Type' dropdown menu (with options: Name, SMILES, InChI, Formula), and a 'Search' button. Below the form is a text box containing the example 'Eg. Aspirin, Alcohol'.

Figura 14 - Pesquisa simples no sistema B3Info


Após o utilizador escolher o identificador, inserir o composto e clicar no botão *Search* a informação é enviada para o servidor para a processar. O processamento é iniciado com a validação da molécula introduzida e de seguida é realizada a pesquisa na base de dados pelo composto introduzido. Se o identificador da molécula introduzida for fórmula molecular será realizada directamente uma pesquisa na tabela *compound* para ver se existe na mesma, caso seja um SMILES ou InChI será obtido o InChIKey correspondente e será efectuada a pesquisa na base de dados. Caso o composto exista na base de dados é recolhida toda a informação associada ao composto nas diversas tabelas e esta é retornada ao utilizador. A Figura 18 é um exemplo da informação apresentada ao utilizador após uma pesquisa pelo composto “aspirin”.

**Blood Brain Barrier Info** Quick Search Structural Search Substructural Search Search By Similarity Download Help B3PP Administration

Molecule Identifier  
aspirin
Type  
Name
Search

Searching for aspirin in the Name category.

**Names**  
2-acetyloxybenzoic acid; 2-Acetoxybenzoic acid; 59-78-2; 11126-35-5; 11126-37-7; 2349-94-2; 26914-13-6; 38291-69-6; NCGC0009977-04; KEGG01\_002272  
**Canonical SMILES**  
CC(=O)Oc1ccccc1C(=O)O  
**InChI**  
InChI=1S/C9H8O4/c1-6(10)/13-8-5-3-2-4-7(8)/11)12h2-5H,1H3,(H,11,12)  
**InChI Key**  
BSYNRYMUTXBVISQ-UNFFFAOYSA-N  
**CAS Registry Number**  
59-78-2


**Molecular Formula**  
C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>  
**Molecular Weight**  
180.1574  
**External Links**  
ChemSpider  
DrugBank  
Wikipedia  
ChEBI  
Check With B3PP

**Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms**  
**BBB Classification**  
(N/A)  
**LogBB**  
BBB-  
**Reference All**  
Dongger, Scott and Hofmann, Thomas and Yeh, Joanne  
Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms, 2002, Vol. 9, Pages 649 to 664  
Reference Link

**Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes**  
**BBB Classification**  
(N/A)  
**LogBB**  
BBB+  
**Reference All**  
Zhao, Yuan H. and Abraham, Michael H. and Ibrahim, Adam and Fish, Paul V. and Cole, Susan and Lewis, Mark L. and de Groot, Marcel J. and Reynolds, Derek P.  
Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes, 2007, Vol. 47, Pages 170 to 175  
Reference Link

**Users Comments**  
Edit  
No comments from users were found for this molecule.

Figura 15 - Exemplo de pesquisa simples pelo composto aspirin“ no B<sub>3</sub>Info

### 6.4.3 Pesquisa Estrutural

A funcionalidade de *Structural Search* (Figura 19) permite ao utilizador construir graficamente um composto através da applet java e realizar uma pesquisa por esse composto no sistema.

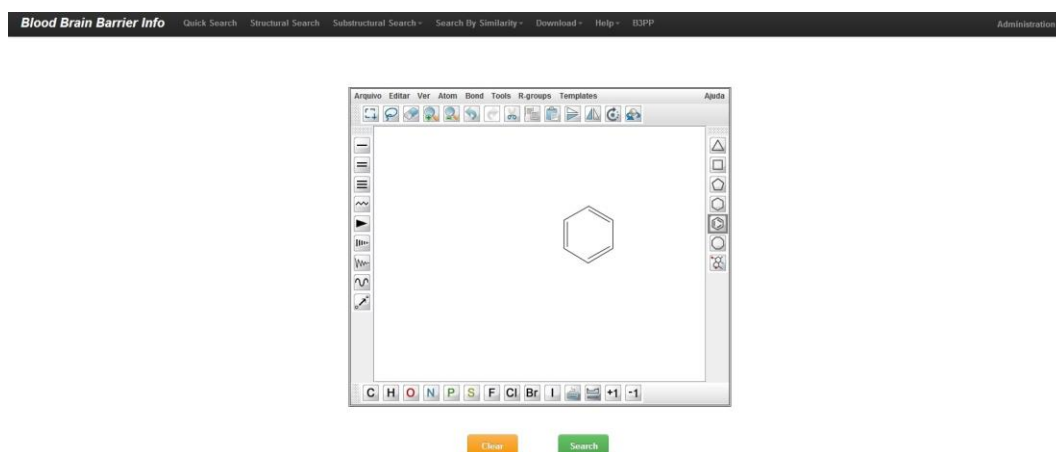


Figura 16 - Exemplo de pesquisa estrutural

O utilizador após construir graficamente o composto e clicar no botão *Search* a applet java converte a estrutura desenhada no SMILES correspondente através do método *getSmiles()* do JChemPaint. Após realizar a operação de conversão o SMILES é enviado para o servidor, é executado no script *Python* que utiliza a biblioteca *Open Babel* para obter o InChIKey correspondente e de seguida é realizada a pesquisa na tabela *compound*. Caso haja correspondência na tabela, é recolhida toda a informação associada ao composto para retornar ao utilizador, caso contrário é enviada a informação de que o composto introduzido não existe no sistema. A Figura 20 representa o retorno da pesquisa no sistema pelo composto desenhado na Figura 19.

The screenshot shows the 'Blood Brain Barrier Info' website interface. At the top, there's a navigation bar with links like 'Quick Search', 'Structural Search', 'Substructural Search', etc. Below this, a green 'Another Search' button is visible. The main content area displays search results for the query 'c1ccccc1'. It includes a chemical structure diagram of a benzene ring. Key information provided includes: Names (Benzene, with various IDs), Canonical SMILES (c1ccccc1), InChI (InChI=1S/C6H6/c1-2-4-5-3-1h1-6H), InChI Key (UHCVQZJYSORH-UHFFFAOYSA-N), CAS Registry Number (27271-55-2), Molecular Formula (C6H6), and Molecular weight (78.1118). There are also links for 'External Links' and 'Check Web Entry'. A reference is cited: 'Effect of Selection of Molecular Descriptors on the Prediction of Blood-Brain Barrier Penetrating and Nonpenetrating Agents by Statistical Learning Methods, 2005, Vol. 45, Pages 1378 to 1384'. At the bottom, there's a 'Users Comments' section with an 'Edit' button and a message stating 'No comments from users were found for this molecule.'

Figura 20 - Exemplo de retorno de pesquisa estrutural

## 6.4.4 Pesquisa Subestrutural

A funcionalidade *Substructural Search* permite ao utilizador realizar pesquisas de compostos com características comuns, tais como átomos, número atómico, tipo de ligações, entre outras. Estas condições podem ser especificadas através de **SMARTS** (Figura 21) ou **graficamente** (Figura 19).

The screenshot shows the 'Blood Brain Barrier Info' website interface for the 'Substructural Search' section. It features a text input field labeled 'SMARTS' with the placeholder '[H][O][H]'. To the right of the input field is a blue 'Search' button. Below the input field, there's an example text: 'Ex: [H][O][H]'. The website's navigation bar is visible at the top.

Figura 21 - Exemplo de uma pesquisa por SMARTS

Após o utilizador inserir o SMARTS e clicar o botão *Search* os dados são enviados para o servidor através da tecnologia javascript. No caso de optar por construir graficamente o composto através da applet java, ao clicar no botão *Search* este é convertido no SMILES correspondente através do método *getSmiles()* do JChemPaint e enviado para o servidor. Após a receção dos dados pelo servidor, o processamento dos dados é igual para ambos os tipos de *input*. Estes são validados e executados no script *Python* que compara o composto introduzido com todos os existentes na base de dados, esta comparação é realizada através do SMILES. De modo a otimizar o desempenho do processo, os SMILES correspondentes aos elementos da base de dados encontram-se armazenados em ficheiro. Após a recolha, esta é organizada e apresentada ao utilizador como mostra a Figura 22.

**Figura 22 - Resultado de uma pesquisa por subestrutura**

[illegible]

Após obter os resultados da pesquisa, o utilizador tem a opção de obter mais informação sobre o composto pretendido. Para tal é necessário clicar no botão *More Info*, que efectua um pedido ao servidor de informação do composto pretendido. Através da tecnologia javascript é inserida na interface a informação associada ao composto pretendido (Figura 24).

**Figura 24 – Exemplo da funcionalidade do botão *More Info***

### 6.4.5 Pesquisa por Similaridade

A funcionalidade *Similar Search* permite ao utilizador realizar pesquisas por compostos semelhantes, através de **SMILES** ou construindo **graficamente** (*Draw Molecule*) o composto a pesquisar. Permite também escolher qual o nível de similaridade pretendido.

Conforme visível nas Figura 25 é necessário a introdução do SMILES e escolher a similaridade pretendida.

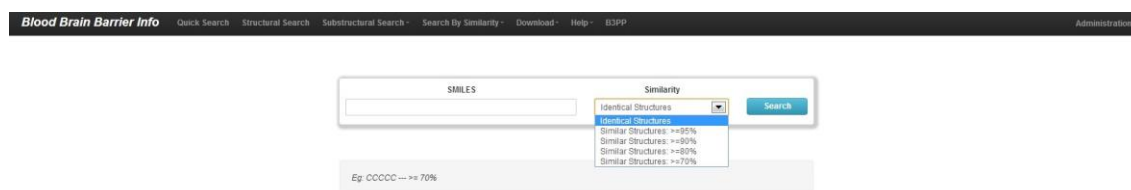


Figura 25 – Pesquisa por similaridade introduzindo um SMILES

Conforme visível na Figura 26 é possível a construção gráfica do composto e escolher o nível de similaridade.

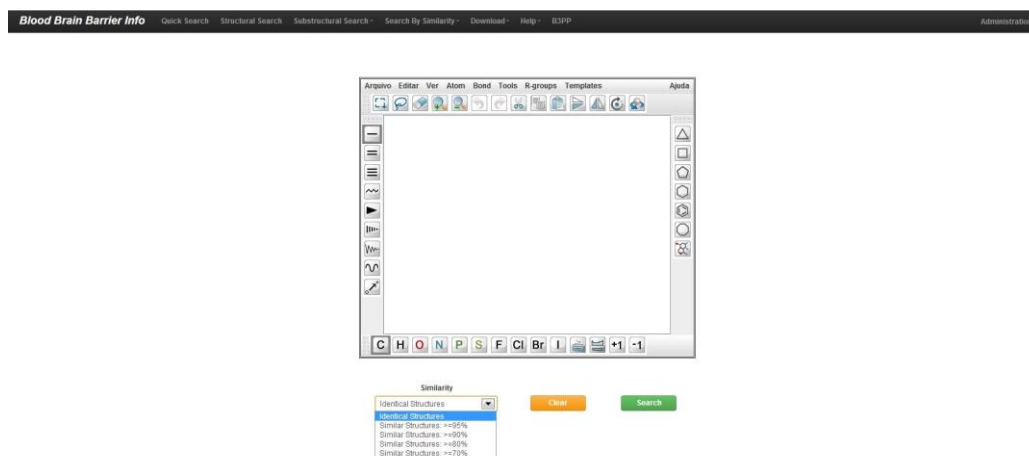


Figura 26 – Pesquisa por similaridade através da applet java

Após a construção gráfica do composto, escolhido o nível de similaridade e clicar no botão *Search*, é obtido o SMILES do composto através da função *getSmiles()* do JChemPaint e o SMILES é enviado para o servidor. No caso de o utilizador optar pela pesquisa por SMILES, após a introdução do mesmo os dados são enviados para o

servidor. Após a recepção dos dados pelo servidor, o processamento é igual para os diferentes tipos de *input*. Estes são validados e de seguida são executados no script *Python* que começa por calcular a sequência de dígitos binários (*fingerprint*) do composto através da biblioteca *Pybel* e de seguida utiliza o coeficiente de Tanimoto para calcular o nível de similaridade entre o composto inserido e os existentes na base de dados, faz a verificação se o valor se encontra de acordo com o seleccionado pelo utilizador na interface. De seguida é pesquisada na base de dados toda a informação relacionada com esses compostos, é organizada e retornada ao utilizador. De modo a melhorar o desempenho do processo, os fingerprints correspondentes aos compostos da base de dados encontram-se armazenados num ficheiro.

Como é visível na Figura 27, a apresentação dos dados ao utilizador segue os mesmos princípios que foram descritos na secção anterior, ou seja, é apresentado o total de dados que cumprem os requisitos, porém só são apresentados na interface os 100 compostos mais semelhantes com o introduzido, existindo a possibilidade de realizar *download* da totalidade dos mesmos através do botão *Download Results*.

The screenshot displays the 'Blood Brain Barrier Info' web application interface. At the top, a navigation bar includes links for 'Quick Search', 'Structural Search', 'Substructural Search', 'Search By Similarity', 'Download', 'Help', 'B3PP', and 'Administration'. The main search area features a 'SMILES' input field with 'CCCCC' entered, a 'Similarity' dropdown menu set to 'Similar Structures: >=70%', and a 'Search' button. Below this, a status bar indicates the search criteria and the number of results: 'You are searching by SMILES CCCCC with similarity >= 70%. Number of compounds found on the database: 8.' An orange 'Download Results' button is positioned to the right. Two compound entries are displayed in a list. Each entry includes the compound name, its InChI string, molecular formula, BBB Classification, and LogBB value. For Hexane, the formula is C<sub>6</sub>H<sub>14</sub>, BBB Classification is BBB+, and LogBB is (N/A). For 3-Methylhexane, the formula is C<sub>7</sub>H<sub>16</sub>, BBB Classification is BBB+, and LogBB is (N/A). Each entry also features a 'More Info' button and a small chemical structure icon.

**Figura 27 – Pesquisa por similaridade através de SMILES**

O utilizador ao clicar no botão *More Info* expande vista sobre o composto, obtendo informação mais completa sobre o mesmo (Figura 24). Os compostos sobre os quais tenha sido requisitada informação mais completa desabilitam o botão *More Info* e activam o botão *Close Info*.

## 6.4.6 Download da Base de Dados

A funcionalidade de download da base de dados permite ao utilizador realizar o download de todos os compostos e informação relacionada com os mesmos para ficheiro para uso próprio. Está disponível o download da informação em ficheiro XML ou ficheiro SDF consoante a preferência do utilizador (Figura 28).



Figura 28 – Menu de escolha do tipo de ficheiro para download

O utilizador ao clicar no botão *Download Database XML File* envia um pedido ao servidor de toda a base de dados e indica que pretende o formato XML. O servidor realiza a pesquisa às tabelas *compound*, *reference*, *compound\_link*, *compound\_name* e *compound\_data\_ref* e processa o ficheiro XML de acordo com *tags* equivalentes aos nomes das tabelas e colunas (Figura 29). Por fim retorna o ficheiro ao utilizador.

```
<?B3Info database version 0.4?>
-<B3Info_database>
-<Molecules>
-<compound>
  <e_id>1</e_id>
  <inchi>
    InChI=1S/C15H24N4S/c20-15(18-13-4-2-1-3-5-13)19-8-6-12(7-9-19)14-10-16-11-17-14/h10-13H,1-9H2,(H,16,17)(H,18,20)
  </inchi>
  <inchi_key>QKDDJDBFONZGBW-UHFFFAOYSA-N</inchi_key>
  <canonical_smiles>S=C(N1CCC(CC1)c1c[nH]cn1)NC1CCCCC1</canonical_smiles>
  <casrn>106243-16-7</casrn>
  <formula>C15H24N4S</formula>
  <mw>292.4429</mw>
</compound>
-<compound>
  <e_id>2</e_id>
  <inchi>
    InChI=1S/C18H18ClNO/c1-20-9-8-13-15-10-12(19)6-7-18(15)21-17-5-3-2-4-14(17)16(13)11-20/h2-7,10,13,16H,8-9,11H2,1H3/t13-,16-/m1/s1
  </inchi>
  <inchi_key>SYLUBQOFGXPQIZ-CJNGLKHVSA-N</inchi_key>
  <canonical_smiles>CN1CC[C@H]2[C@H](C1)c1cccc1Oe1c2cc(Cl)cc1</canonical_smiles>
  <casrn>(N/A)</casrn>
  <formula>C18H18ClNO</formula>
  <mw>299.7946</mw>
</compound>
-<compound>
  <e_id>3</e_id>
  <inchi>
    InChI=1S/C18H24ClF3N4O/c19-15-12-14(18(20,21)22)13-16(23-15)25-10-8-24(9-11-25)5-1-2-6-26-7-3-4-17(26)27/h12-13H,1-11H2
  </inchi>
  <inchi_key>XSQCKDORBNQRDB-UHFFFAOYSA-N</inchi_key>
  <canonical_smiles>O=C1CCCN1CCCCN1CCN(CC1)c1nc(Cl)cc(e1)C(F)(F)F</canonical_smiles>
```

Figura 29 – Ficheiro XML que contém toda a informação da base de dados





De forma a facilitar a análise da estrutura do ficheiro XML e do modelo relacional, é fornecido o *XML Schema* do ficheiro XML. Este pode ser analisado directamente na página web da funcionalidade (Figura ZZ). Porém existe também a funcionalidade de download do *XML Schema* em ficheiro XML, para tal é necessário o utilizador clicar no botão *Download XML Schema* e é enviado um pedido ao servidor a indicar o pedido. Este inicia o processamento com uma consulta às tabelas da base de dados e colunas das mesmas de modo a gerar o ficheiro XML com as *tags* correctas. Após o processamento é retornado o ficheiro ao utilizador.

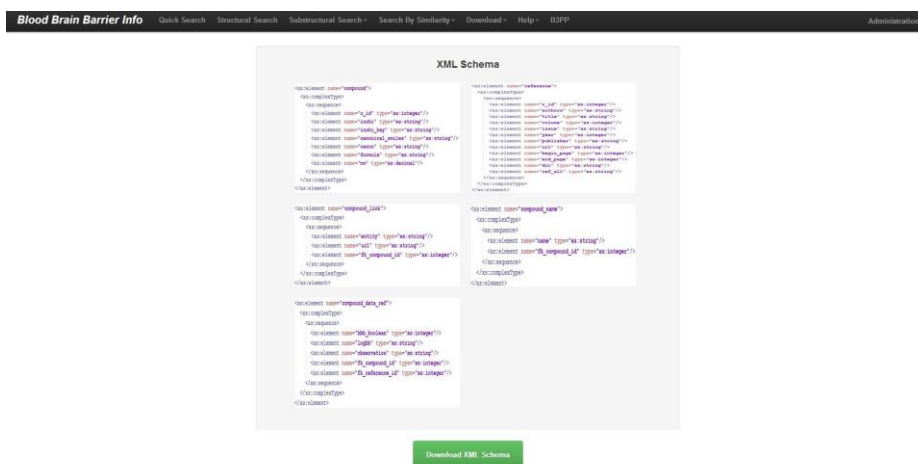


Figura 31 – XML Schema

## 6.4.7 Secção de Ajuda

A secção **B3Info Help** permite ao utilizador esclarecer as dúvidas e ver exemplos sobre todas as funcionalidades disponíveis no menu superior do sistema.

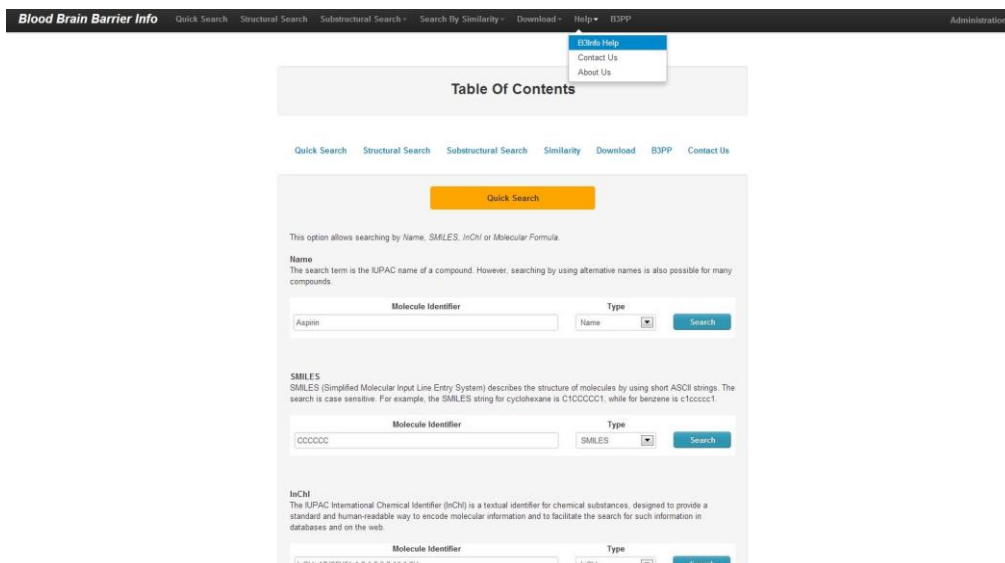


Figura 32 – B3Info Help

Caso o utilizador sinta a necessidade de contactar a administração do sistema para, tem ao seu dispor a funcionalidade **Contact Us** (Figura XX) que apresenta um formulário de fácil preenchimento. Os dados são enviados para o servidor, e este inicia o processamento dos mesmos com a validação do email e se todos os campos estão preenchidos, de seguida realiza uma pesquisa á tabela *user* para obter os emails de todos os administradores e envia um mail a todos com o texto inserido pelo utilizador. Após o envio do email para aos administradores é enviada uma mensagem de sucesso/insucesso da operação ao utilizador.

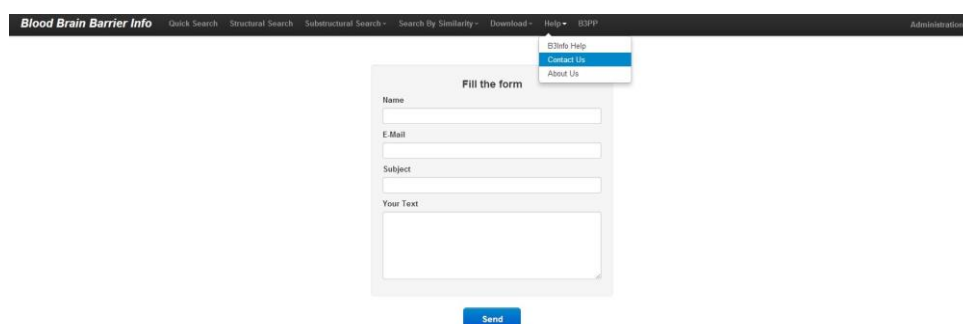


Figura 33 – Funcionalidade de contacto com os administradores

Se o utilizador tiver curiosidade de obter mais informação sobre a equipa de desenvolvimento do sistema *B3Info*, pode aceder á vista **About Us** que contém a listagem das pessoas envolvidas no projecto e a hiperligação para a página pessoal das mesmas (Figura 34).

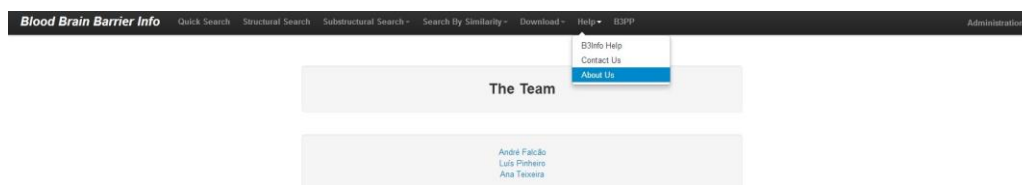


Figura 34 – Constituição da equipa de desenvolvimento do B<sub>3</sub>Info

### 6.4.8 Comunicação com o *B<sub>3</sub>PP*

A qualquer momento o utilizador pode aceder à aplicação web *B<sub>3</sub>PP* pois a funcionalidade de interacção com a aplicação *B<sub>3</sub>PP* encontra-se disponível no menu superior (Figura 35). Para tal é necessário clicar sobre o ícone *B<sub>3</sub>PP* e é aberto num novo separador a interface da aplicação.



Figura 35 – Menu superior do sistema B3Info

Quando é realizada uma pesquisa por um composto e o resultado é apresentado na interface, o utilizador pode comparar o resultado obtido no *B3Info* com o modelo de predição da permeabilidade de um composto na BHE disponibilizado pelo *B3PP*. Para tal, é necessário efectuar um clique no botão *Check With B3PP* disponibilizado na interface (Figura 36).



Figura 36 – Resultado da previsão do composto *Aspirin* no *B3PP*

Após efectuado clique no botão *Check With B3PP* é enviado para o servidor o SMILES do composto pesquisado. Este valida o SMILES recebido e prepara os argumentos necessários para invocar o URL do *Web Service* disponibilizado pela aplicação *B3PP* através da biblioteca cURL do PHP. Após obter a resposta por parte do *Web Service*, o servidor organiza os dados e apresenta-os ao utilizador (Figura 36). Para fornecer feedback, após o utilizador clicar no botão é inserido na interface o estado de *loading* enquanto o processo de executa.

### 6.4.9 Área de administração

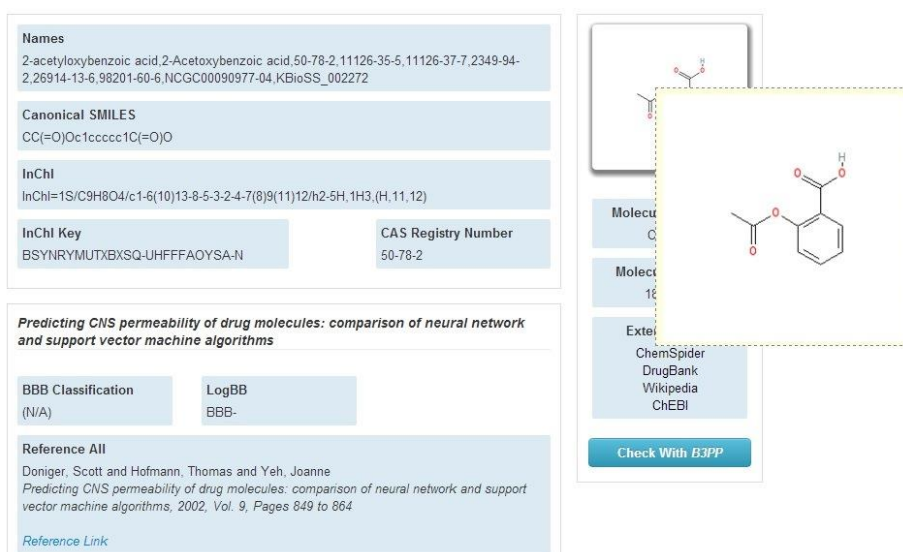
Para aceder às funcionalidades de gestão do sistema disponíveis no backend o utilizador apenas tem que clicar na hiperligação disponibilizada no menu superior. Após efectuar o clique é aberto um novo separador com a interface de login para acesso ao ao mesmo.



Figura 37 – Hiperligação para acesso ao backend do sistema B3Info

### 6.4.10 Funcionalidades Comuns

Após o utilizador efectuar uma pesquisa de um componente, é-lhe apresentada a interface com o resultado da mesma. Se mover o rato para cima da imagem correspondente ao composto, é efectuado zoom sobre a mesma (Figura XX). Esta funcionalidade foi desenvolvida através da tecnologia CSS.



The screenshot displays the B3Info interface for a chemical compound. On the left, a panel contains the following information:

- Names:** 2-acetyloxybenzoic acid, 2-Acetoxybenzoic acid, 50-78-2, 11126-35-5, 11126-37-7, 2349-94-2, 26914-13-6, 98201-60-6, NCGC00090977-04, KBioSS\_002272
- Canonical SMILES:** CC(=O)Oc1ccccc1C(=O)O
- InChI:** InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)
- InChI Key:** BSYNRYMUTXBXSQ-UHFFFAOYSA-N
- CAS Registry Number:** 50-78-2

Below this panel, a section titled "Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms" includes:

- BBB Classification:** (N/A)
- LogBB:** BBB-
- Reference All:** Doniger, Scott and Hofmann, Thomas and Yeh, Joanne. Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms, 2002, Vol. 9, Pages 849 to 864.
- Reference Link:** (a blue hyperlink)

On the right side of the interface, there is a chemical structure of 2-acetoxybenzoic acid. A dashed yellow box highlights the structure, indicating a zoomed-in view. Below the structure, there are links to external resources: ChemSpider, DrugBank, Wikipedia, and ChEBI. A button labeled "Check With B3PP" is also visible.

Figura 38 – Zoom sobre a imagem do composto

De modo a complementar a informação disponibilizada pelo sistema B3Info, são disponibilizadas ligadas externas ao sistema, ou seja, hiperligações para outros sistemas de informação que contêm informações relevantes sobre o composto pesquisado (Figura 39). O utilizador ao clicar sobre o sistema pretendido é aberto um novo separador com a informação do composto no sistema destino.



Figura 39 – Ligações externas associadas ao composto *Aspirin*

Caso este pretenda também consultar a publicação onde são anunciados os valores de penetração na BHE por parte do composto pesquisado, existe também a hiperligação para a mesma através do *Reference link*.

De modo a possibilitar mais interacção entre o utilizador e o sistema e de enriquecer a informação disponibilizada pelo mesmo, foi implementada a funcionalidade de inserir comentários na molécula pesquisada (Figura 40). Inicialmente o área de texto dedicada aos comentário encontra-se em modo de leitura, após o utilizador efectuar um clique no botão *Edit* passa para modo de escrita e aparece o botão *Save* para guardar os comentários inseridos. Após clicar no botão *Save* o texto é enviado para o servidor e este insere o texto na tabela *compound\_user\_comments* e associa este comentário com o composto em indicado.



Figura 40 – Funcionalidade de inserir comentários sobre a molécula pesquisada

## 6.5 Backend

O backend permite aos administradores do sistema B<sub>3</sub>Info gerir todos os dados armazenados na base de dados, ou seja, adicionar, editar e eliminar moléculas, referências, relações molécula-referência e utilizadores. Permite também a actualização do sistema e ficheiros necessários ao seu correcto funcionamento.

### 6.5.1 Interface Backend B3Info

Interface (Figura 41) desenvolvida a pensar no administrador do sistema, com todas as funcionalidades necessárias á gestão do mesmo acessíveis de forma directa e intuitiva para que o tempo de aprendizagem necessário seja reduzido.

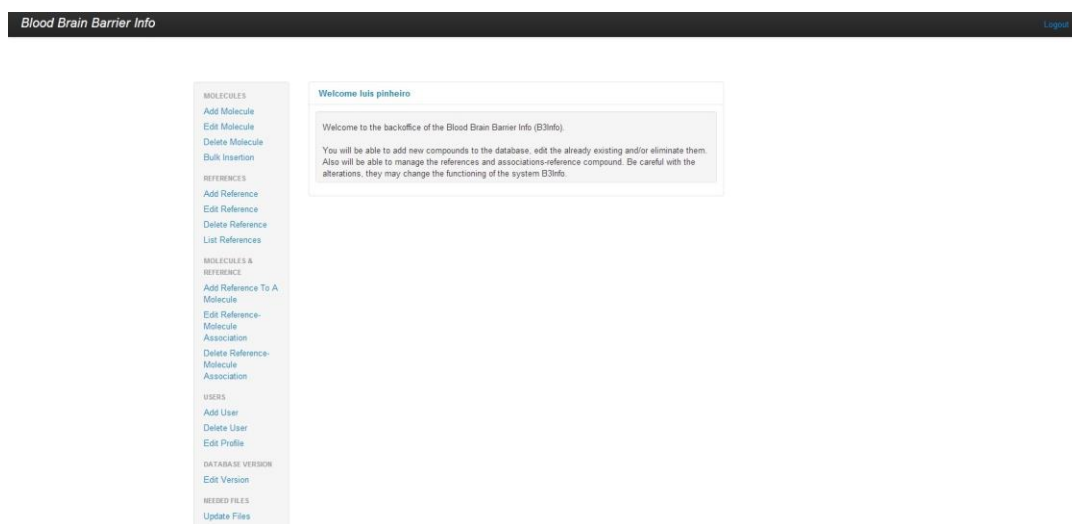
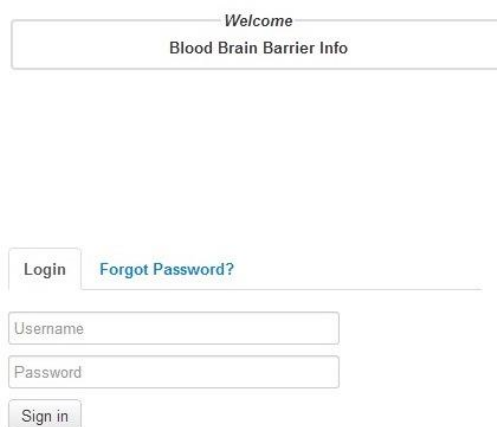


Figura 41 - Interface Backend do sistema B3Info

Esta foi desenvolvida recorrendo às tecnologias HTML e CSS para definição e preenchimento do layout, menus, botões, cores, estilos de letra e animações. Foi também utilizada a tecnologia javascript para modificação dos elementos HTML, ou seja, adição de estados de *processing* para fornecer feedback ao utilizador e posterior introdução na página web dos resultados da operação.

## 6.5.2 Autenticação

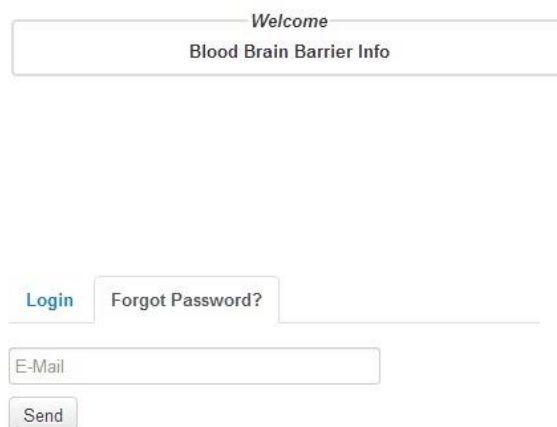
Para ter acesso ao backend é necessário efectuar o login no sistema através da introdução do *username* e *password* no como é visível na Figura 42. Após o preenchimento dos campos necessários, estes são enviados para o servidor que vai á tabela *user* validar os dados. Caso os dados sejam válidos, o utilizador é encaminhado para a pagina inicial do backend, caso contrário é informado que os dados não são válidos.



The screenshot shows a web interface for the 'Blood Brain Barrier Info' system. At the top, there is a header box with the text 'Welcome' and 'Blood Brain Barrier Info'. Below this, there is a login section. It features a horizontal bar with two buttons: 'Login' and 'Forgot Password?'. Underneath this bar are two input fields: 'Username' and 'Password'. At the bottom of the login section is a 'Sign in' button.

Figura 42 – Menu de login do backend do sistema B<sub>3</sub>Info

Existe também a possibilidade de o utilizador recuperar os dados de acesso, para tal é necessário introduzir o email com o qual está registado no sistema (Figura 43). Após a inserção do email e clicar em *Send* o servidor recebe o email introduzido, valida o mesmo e de seguida verifica se este existe na tabela *user*. Caso haja correspondência é enviado um mail com as credenciais de acesso para o email especificado, caso contrario o utilizador é informado que o email não existe no sistema.



The screenshot shows a web interface for the 'Blood Brain Barrier Info' system. At the top, there is a header box with the text 'Welcome' and 'Blood Brain Barrier Info'. Below this, there is a password recovery section. It features a horizontal bar with two buttons: 'Login' and 'Forgot Password?'. Underneath this bar is a single input field labeled 'E-Mail'. At the bottom of the section is a 'Send' button.

Figura 43 - Funcionalidade de recuperar dados de acesso ao backend



### 6.5.3 Gestão de Moléculas

A gestão de moléculas permite ao administrador do sistema adicionar apenas uma molécula ou em massa, editar e eliminar moléculas da base de dados.

#### Adicionar Molécula

A funcionalidade *Add Molecule* permite inserir uma molécula na base de dados. Para tal o administrador do sistema começa por inserir o identificador da molécula (SMILES ou InChI), os nomes e o casrn da mesma. De seguida insere os dados relacionados com a penetração na BHE, ou seja, a classe BBB a que pertence (positivo, negativo ou indeterminado), o LogBB, as observações e a publicação onde a informação de permeabilidade foi validada. Opcionalmente pode inserir ligações para sistemas externos (Figura 44).

The screenshot shows a web interface for adding a molecule. On the left is a sidebar menu with categories: MOLECULES (Add, Edit, Delete Molecule, Bulk Insertion), REFERENCES (Add, Edit, Delete Reference, List References), MOLECULES & REFERENCE (Add Reference To A Molecule, Edit Reference-Molecule Association, Delete Reference-Molecule Association), USERS (Add User, Delete User, Edit Profile), DATABASE VERSION (Edit Version), and NEEDED FILES (Update Files). The main form is titled 'Add Molecule' and contains the following sections:

- Structural Data:** Includes a 'Molecule Format' section with radio buttons for 'InChI' and 'SMILES'. Below it is a text input for 'Names (Optional) (Separated by commas)' and another for 'CASRN (Optional)'.
- BBB Penetration Data:** Includes a 'Select BBB Penetration Class' section with radio buttons for 'BBB+', 'BBB-', and 'Unspecified'. Below it is a text input for 'LogBB'.
- Reference:** Includes a 'Select the reference to the compound' section with a dropdown menu showing 'Predicting CNS permeability of drug'.
- External Links:** Includes input fields for 'DrugBank', 'PubChem', 'Wikipedia', 'ChEBI', and 'ChemSpider'.
- Observations:** Includes a large text area for notes.

A blue 'Save' button is located at the bottom center of the form.

Figura 44 – Funcionalidade de inserir molécula na base de dados

O servidor recebe os dados enviados pelo utilizador e começa o seu processamento por verificar se o composto já existe na base de dados, se já existir verifica se existe a relação composto-referência. Se esta também já existir informa o utilizador que os dados introduzidos já existem no sistema, caso contrário inicia o processamento dos mesmos e posterior inserção na base de dados. No final informa o utilizador que os dados foram inseridos com sucesso.

## Edição de Moléculas

A funcionalidade de edição de molécula permite ao administrador editar os dados associados a um composto. De modo a facilitar a pesquisa de quais as moléculas no sistema, o utilizador é assistido na escolha do composto pela funcionalidade de *autocomplete*.

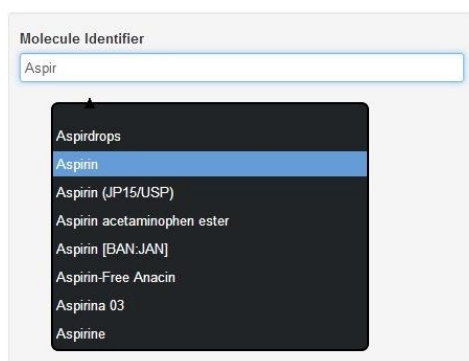


Figura 45 – Funcionalidade de *autocomplete*

Após seleccionar o composto a editar, é carregada a vista com as opções que permite editar os dados do mesmo (Figura 46).

Figura 46 – Funcionalidade de editar um composto

Como se pode verificar na Figura 46 os dados estruturais não é possível alterar pois são recolhidos através da execução de scripts *Python* e são únicos para a molécula em causa. Em relação aos outros atributos do composto, são editáveis. Após o utilizador editar o atributo pretendido e guardar, os dados que foram editados são enviados para o servidor, são processados e inseridos na base de dados.

## Eliminação de Moléculas

A funcionalidade *Delete Molecules* permite eliminar um composto da base de dados (Figura 47). Ao eliminar o composto, serão também eliminados todos os dados associados ao composto em causa.

MOLECULES  
Add Molecule  
Edit Molecule  
Delete Molecule  
Bulk Insertion

REFERENCES  
Add Reference  
Edit Reference  
Delete Reference  
List References

MOLECULES & REFERENCE  
Add Reference To A Molecule  
Edit Reference-Molecule Association  
Delete Reference-Molecule Association

USERS  
Add User  
Delete User  
Edit Profile

DATABASE VERSION  
Edit Version

NEEDED FILES  
Update Files

**Molecule To Remove**

Molecule Identifier: Aspirin

Type of compound: Name

**Molecule Data**

**Names**  
2-acetyloxybenzoic acid 2-Acetoxybenzoic acid.50-78-2,11126-35-5,11126-37-7,2349-94-2,26914-13-6,98201-60-6,NCGC00090977-04,KBioSS\_002272

**Canonical SMILES**  
CC(=O)Oc1ccccc1C(=O)O

**InChI**  
InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)

**Formula**  
C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>

**Molecular Weight**  
180.1574

**CAS-RN**  
50-78-2

**Delete**

Figura 47 – Eliminar composto

## Inserção em Massa

A funcionalidade *Bulk Insertion* permite inserir compostos em bruto na base de dados. Para tal é necessário indicar qual o identificador dos compostos, a sequência com que os dados estão dispostos e qual a referencia a que ficam associados. Existe a opção de inserir a partir de um ficheiro com extensão *.txt* ou através da caixa de texto disponível na vista (Figura 48).

**MOLECULES**  
 Add Molecule  
 Edit Molecule  
 Delete Molecule  
 Bulk Insertion

**REFERENCES**  
 Add Reference  
 Edit Reference  
 Delete Reference  
 List References

**MOLECULES & REFERENCE**  
 Add Reference To A Molecule  
 Edit Reference-Molecule Association  
 Delete Reference-Molecule Association

**USERS**  
 Add User  
 Delete User  
 Edit Profile

**DATABASE VERSION**  
 Edit Version

**NEEDED FILES**  
 Update Files

**Common Properties**

**Molecule Identifier**  
☐ InChI  
☐ SMILES

**Data Sequence**  
☐ ID BBB+/- Observations  
☐ ID BBB+/- LogBB Observations  
☐ ID LogBB Observations

**Select the reference**  
 Predicting CNS permeability

**Input Type**  
☒ Text Box ☐ File

**Molecules Sequences**

**Validation** **Insert All**

**Figura 48 – Inserção em massa**

Após inserir os dados é necessário efectuar a validação dos mesmos através do botão *Validation*, que envia os dados para o servidor, e este vai verificar se estão de acordo com a sequência indicada. O retorno do servidor indica quais os dados que estão de acordo com a sequência e quais os que necessitam de correcção (Figura 49).

**Common Properties**

**Molecule Identifier**  
☐ InChI  
☒ SMILES

**Data Sequence**  
☒ ID BBB+/- Observations  
☐ ID BBB+/- LogBB Observations  
☐ ID LogBB Observations

**Select the reference**  
 Predicting CNS permeability

**Examples**

Identifier	BBB+/-	Observations
CCCC	p	obs

**Input Type**  
☒ Text Box ☐ File

**Molecules Sequences**

```
CCC p observation1
CCCC p observation2
CCCC test observation3
```

**The following lines have errors:**

```
CCCC test observation3
```

**The following lines are OK:**

```
CCC p observation1
CCCC p observation2
```

**Validation** **Insert All**

**Figura 49 – Valida sequencia**

Quando todas as linhas estiverem na sequência correcta é que o botão *Insert All* fica activo. Após ficar activo e o administrador enviar os dados para o servidor, este

inicia o processamento dos mesmos. Separa-os linha a linha, e de seguida separa cada linha consoante a sequencia indicada e inicia o processamento de cada composto para inserir a base de dados.

### 6.5.4 Gestão de Referências

Esta funcionalidade permite a inserção, edição, eliminação e consulta das referências bibliográficas existentes na base de dados.

#### Adicionar Referência

A funcionalidade *Add Reference* permite a inserção de uma referência bibliográfica na base de dados. Para tal é necessário preencher todos os campos visíveis na Figura 50. Após o envio dos dados para o servidor, este verifica se a referencia já existe na base de dados, se existe informa o administrador, caso contrário insere a referência.

The image shows a web application interface for adding a reference. On the left is a sidebar menu with the following categories and links:

- MOLECULES
  - Add Molecule
  - Edit Molecule
  - Delete Molecule
  - Bulk Insertion
- REFERENCES
  - Add Reference
  - Edit Reference
  - Delete Reference
  - List References
- MOLECULES & REFERENCE
  - Add Reference To A Molecule
  - Edit Reference-Molecule Association
  - Delete Reference-Molecule Association
- USERS
  - Add User
  - Delete User
  - Edit Profile
- DATABASE VERSION
  - Edit Version
- NEEDED FILES
  - Update Files

The main content area is titled 'Reference' and contains a form with the following fields:

- Authors
- Title
- Volume
- Year
- Issue
- Publisher
- URL
- Begin Page
- End Page
- DOI (Document Object Identifier)

Below the form is a blue button labeled 'Save Reference'.

**Figura 50 – dicionar referência bibliográfica**

## Edição de Referência

A funcionalidade *Edit Reference* permite a edição de uma referência bibliográfica existente na base de dados (Figura 51). O Administrador selecciona qual a referência que pretende editar, através do Ajax é realizada uma consulta á base de dados pela referência e são obtidos os dados. O administrador altera os dados pretendidos e envia-os para o servidor, e este actualiza na base de dados.

The screenshot displays the 'Edit Reference' web interface. On the left is a sidebar with navigation links categorized under MOLECULES, REFERENCES, MOLECULES & REFERENCE, USERS, DATABASE VERSION, and NEEDED FILES. The top bar features a dropdown menu labeled 'Select the reference to edit' with the selected item 'QSAR Modeling of the Blood-Brain Barrier Per'. The main form area, titled 'Edit Reference Data', contains several input fields: Authors (Zhang, Liying and Zhu, Hao and Oprea, Tudor i), Title (QSAR Modeling of the Blood-Brain Barrier Per), Volume (25), Issue (8), Year (2008), Publisher (Springer Netherlands), URL (http://dx.doi.org/10.1007/s11095-008-9609-0), Begin Page (1902), End Page (1914), DOI (10.1007/s11095-008-9609-0), and Ref All (QSAR Modeling of the Blood-Brain Barrier Per). A 'Save Changes' button is located at the bottom right of the form.

Figura 51 - Editar uma referencia bibliográfica

## Remoção de Referência

A funcionalidade *Delete Reference* permite eliminar uma referência bibliográfica da base e dados. O administrador selecciona qual a referência que pretende eliminar (Figura 52), envia a informação para o servidor e este elimina a mesma da base de dados. Todas as associações à referência em causa são também eliminadas.

The screenshot displays the 'Delete Reference' web interface. On the left is a sidebar with navigation links categorized under MOLECULES, REFERENCES, MOLECULES & REFERENCE, USERS, DATABASE VERSION, and NEEDED FILES. The top bar features a dropdown menu labeled 'Select the reference to remove' with the selected item 'Predicting CNS permeability of drug molecule'. The main form area, titled 'Delete Reference', contains several input fields: Authors (Domger, Scott and Hofmann, Thomas and Yeh, Joanne), Title (Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms), Issue, Publisher (Journal of computational biology), URL (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\_u), Volume (9), Year (2002), DOI (10.1089/10665270260518317), Begin Page (849), and End Page (864). A 'Delete Reference' button is located at the bottom right of the form.

Figura 52 – Eliminar referência

## Listagem de Referências

A funcionalidade *List References* permite ao administrador consultar todas as referências existentes na base de dados (Figura 53). Realiza uma consulta á base de dados por todas as referências e estas são retornadas ao administrador.

<b>MOLECULES</b> <a href="#">Add Molecule</a> <a href="#">Edit Molecule</a> <a href="#">Delete Molecule</a> <a href="#">Bulk Insertion</a>	<b>List Of All References</b>
<b>REFERENCES</b> <a href="#">Add Reference</a> <a href="#">Edit Reference</a> <a href="#">Delete Reference</a> <a href="#">List References</a>	
<b>MOLECULES &amp; REFERENCE</b> <a href="#">Add Reference To A Molecule</a> <a href="#">Edit Reference-Molecule Association</a> <a href="#">Delete Reference-Molecule Association</a>	
<b>USERS</b> <a href="#">Add User</a> <a href="#">Delete User</a> <a href="#">Edit Profile</a>	
<b>DATABASE VERSION</b> <a href="#">Edit Version</a>	
<b>NEEDED FILES</b> <a href="#">Update Files</a>	

Title	Authors	Volume	Year
Predicting CNS permeability of drug molecules: comparison of neural network and support vector machine algorithms	Doniger, Scott and Hofmann, Thomas and Yeh, Joanne	9	2002
Effect of Selection of Molecular Descriptors on the Prediction of Blood-Brain Barrier Penetrating and Nonpenetrating Agents by Statistical Learning Methods	Li, Hu and Yap, Chun Wei and Ung, Choong Yong and Xue, Ying and Cao, Zhi Wei and Chen, Yu Zong	45	2005
QSAR Modeling of the Blood-Brain Barrier Permeability for Diverse Organic Compounds	Zhang, Liying and Zhu, Hao and Oprea, Tudor and Golbraikh, Alexander and Tropsha, Alexander	25	2008
Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes	Zhao, Yuan H. and Abraham, Michael H. and Ibrahim, Adam and Fish, Paul V. and Cole, Susan and Lewis, Mark L. and de Groot, Marcel J. and Reynolds, Derek P.	47	2007
Computational Prediction of Blood-Brain Barrier Permeability Using Decision Tree Induction	Claudia Suenderhauf, Felix Hammann and Jörg Huwyler	17	2012

Figura 53 – Consulta lista de referências

## 6.5.5 Gestão da Associação Composto-Referencia

### Adicionar Referencia a um Composto

A funcionalidade *Add Reference To A Molecule* permite a criação da associação molécula-referência. O Administrador selecciona a molécula através da funcionalidade *autocomplete* descrita anteriormente, de seguida selecciona a referência, a classe de penetração, o valor do *LogBB* e opcionalmente insere observações. De seguida os dados

são enviados para o servidor, este verifica se associação em causa já existe na base de dados (Figura 54). Se já existe, informa o utilizador, caso contrário insere a informação na base de dados.

The screenshot shows a web interface for managing molecule references. On the left is a sidebar menu with options like 'Add Molecule', 'Edit Molecule', 'Delete Molecule', 'Bulk Insertion', 'Add Reference', 'Edit Reference', 'Delete Reference', 'List References', 'Add Reference To A Molecule', 'Edit Reference-Molecule Association', 'Delete Reference-Molecule Association', 'Add User', 'Delete User', 'Edit Profile', 'Edit Version', and 'Update Files'. The main form is titled 'Molecule' and contains the following sections:

- Molecule:** A text input for 'Molecule Identifier' and a dropdown menu for 'Type of compound' with 'Name' selected.
- Molecule & Reference Data:**
  - Names:** A text area containing chemical names and identifiers: '2-Chloro-1-(difluoromethoxy)-1,1,2-trifluoroethane, 2-chloro-1-(difluoromethoxy)-1,1,2-trifluoroethane, 13838-16-9, EINECS 237-553-4, Efrane, Enflurane [Anaesthetics, volatile], Enflurane [USAN, BAN, INN, JAN], C07516, Enflurane, Ethane, 2-chloro-1-(difluoromethoxy)-1,1,2-trifluoro-'. A chemical structure diagram is shown to the right.
  - InChI:** A text input containing 'InChI=1S/C3H2CF5O1c4-1(5)(3(8,9))10-2(6)7/1-2H'.
  - BBB Classification:** A dropdown menu showing 'BBB+'.
  - LogBB:** A text input showing '(N/A)'.
  - Observations:** A text area showing '(N/A)'.
  - Reference:** A text area containing a citation: 'Li, Hu and Yap, Chun Wei and Ung, Choong Yong and Xue, Ying and Cao, Zhi Wei and Chen, Yu Zong, Effect of Selection of Molecular Descriptors on the Prediction of Blood-Brain Barrier Penetrating and Nonpenetrating Agents by Statistical Learning Methods, 2005, Vol. 45, Pages 1376 to 1384'.
- New Association Data:**
  - Select the reference:** A dropdown menu showing 'Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fra...'.
  - Select BBB Penetration Class:** Radio buttons for 'BBB+' (selected), 'BBB-', and 'Unspecified'.
  - LogBB:** A text input.
  - Observations (Optional):** A large text area.

At the bottom of the form is a blue button labeled 'Save Molecule Reference Association'.

Figura 54 – Adicionar uma associação referencia-molécula

### Editar associação composto-referencia

A funcionalidade *Edit Reference-Molecule Association* permite ao administrador editar uma associação molécula-referência existente na base de dados. Através da funcionalidade *autocomplete* o administrador selecciona qual o composto para editar, de seguida o sistema apresenta todas as associações que esse composto possui (Figura 55). O administrador edita os campos que pretende e envia os dados para o servidor. Este valida os dados e insere-os na base de dados.



Figura 55 – Edição da associação molécula-referência

## Remoção de associação molécula-referencia

A funcionalidade *Delete Reference-Molecule Association* permite ao administrador eliminar uma associação referencia-molécula. Este selecciona qual a molécula que pretende, o sistema retorna todas as associações que o composto possui (Figura 56). De seguida o administrador só necessita de clicar no botão *Delete*, envia a ordem para os servidor e este elimina a molécula da base de dados. Devido á utilização da tecnologia Ajax a associação eliminada desaparece da vista sem ser necessário recarregar novamente a página.

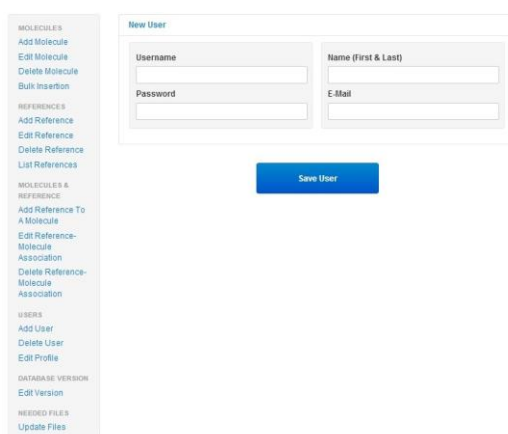
Figura 56 – Eliminar associação Referencia-Molécula

## 6.5.6 Gestão de Utilizadores

Esta secção permite a criação, edição e eliminação de novos administradores do sistema.

### Adicionar Utilizador

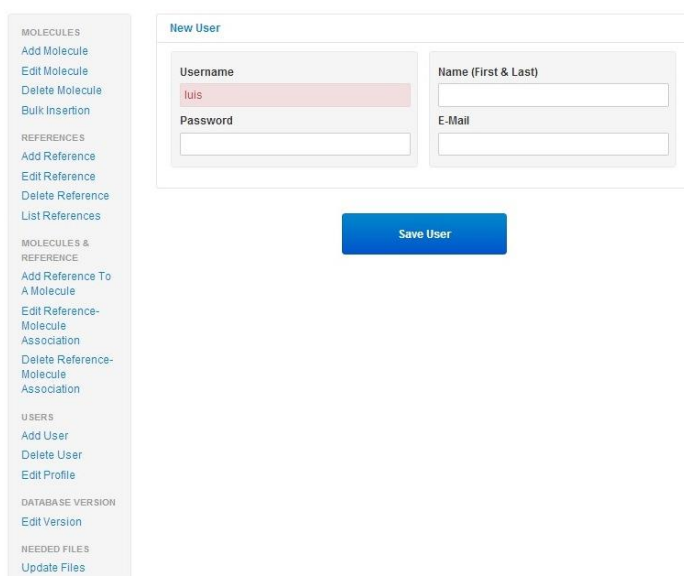
A funcionalidade *Add User* permite adicionar um novo administrador ao sistema. Os campos representados na Figura 57 são de preenchimento obrigatório. Após o seu preenchimento são enviados para o servidor e inseridos na base de dados.



The screenshot shows the 'New User' form. On the left is a sidebar menu with categories: MOLECULES, REFERENCES, MOLECULES & REFERENCE, USERS, DATABASE VERSION, and NEEDED FILES. The 'New User' form has four input fields: Username, Password, Name (First & Last), and E-Mail. A blue 'Save User' button is at the bottom.

Figura 57 – Criação de novo administrador

Quando os dados são enviado para o servidor, não existe e verificação se o administrador já existe na base de dados ou não, pois essa validação é realizada através do Ajax enquanto se preenchem o campos *username* e *email* que são únicos. No caso de algum dos valores contidos nestes campos existir na base de dados , o campo em causa fica com cor de fundo vermelha (Figura 58).



This screenshot is similar to Figure 57, but the 'Username' input field has a red background, indicating a validation error because the username already exists in the database. The 'Save User' button remains at the bottom.

Figura 58 - Administrador já existente no sistema

## Eliminar Administrador

A funcionalidade *Delete User* permite eliminar um administrador do sistema (Figura 59). É necessário escolher qual o administrador que se pretende eliminar e enviar essa informação para o servidor. Este elimina o utilizador indicado.

MOLECULES  
Add Molecule  
Edit Molecule  
Delete Molecule  
Bulk Insertion

REFERENCES  
Add Reference  
Edit Reference  
Delete Reference  
List References

MOLECULES & REFERENCE  
Add Reference To A Molecule  
Edit Reference-Molecule Association  
Delete Reference-Molecule Association

USERS  
Add User  
Delete User  
Edit Profile

DATABASE VERSION  
Edit Version

NEEDED FILES  
Update Files

Select the user to remove  
lpinheiro

User Data

Username  
lpinheiro

Name (First & Last)  
Luís Pinheiro

Password  
\*\*\*\*\*

E-Mail  
lpinheiro@iasige.di.fc.ul.pt

Remove User

Figura 59 – Remover administrador do sistema

## Editar Perfil

A funcionalidade *Edit Profile* permite ao administrador definir uma nova password ou um novo email (Figura 60). Após o preenchimento dos campos, estes são enviados para o servidor, que actualiza a base de dados com os novos valores.

MOLECULES  
Add Molecule  
Edit Molecule  
Delete Molecule  
Bulk Insertion

REFERENCES  
Add Reference  
Edit Reference  
Delete Reference  
List References

MOLECULES & REFERENCE  
Add Reference To A Molecule  
Edit Reference-Molecule Association  
Delete Reference-Molecule Association

USERS  
Add User  
Delete User  
Edit Profile

DATABASE VERSION  
Edit Version

NEEDED FILES  
Update Files

User Data Edit

New Password

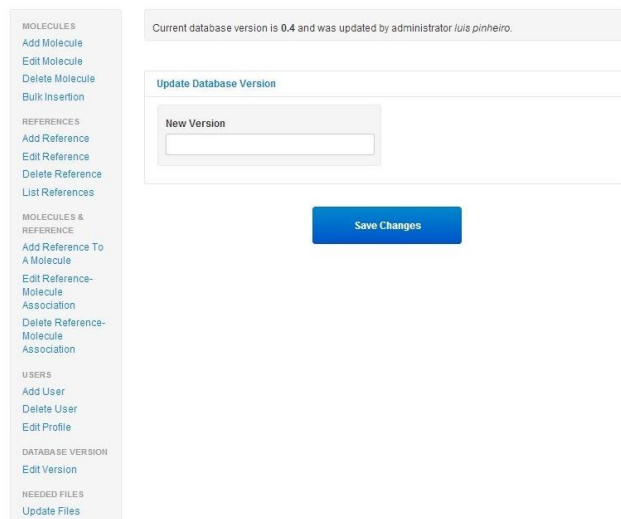
New E-Mail

Save Changes

Figura 60 – Edição do perfil

### 6.5.7 Actualizar versão da base de dados

Esta funcionalidade permite actualizar a versão da base de dados. É importante que sempre que haja alterações na base de dados este valor também seja actualizado de modo a manter a coerência entre as modificações de dados e a versão correspondente.



The screenshot shows a web application interface for updating the database version. On the left is a vertical sidebar menu with categories: MOLECULES (Add, Edit, Delete, Bulk Insertion), REFERENCES (Add, Edit, Delete, List), MOLECULES & REFERENCE (Add Reference To A Molecule, Edit Reference-Molecule Association, Delete Reference-Molecule Association), USERS (Add, Delete, Edit Profile), DATABASE VERSION (Edit Version), and NEEDED FILES (Update Files). The 'Edit Version' option is highlighted. The main content area has a header stating 'Current database version is 0.4 and was updated by administrator luis pinheiro.' Below this is a section titled 'Update Database Version' containing a 'New Version' label and a text input field. At the bottom of this section is a blue 'Save Changes' button.

Figura 61 – Actualizar versão da base de dados

### 6.5.8 Actualizar ficheiros auxiliares

A funcionalidade *Update files* permite actualizar os ficheiros necessários ao melhor funcionamento do sistema (Figura 62). São 3 os ficheiros que é necessário actualizar e cada um deles contém todos os compostos armazenados na base de dados, cada um no seu formato próprio consoante a finalidade. A escolha de ficheiros deveu-se a questões de performance pois revelou-se mais rápido o acesso a ficheiro do que a leitura contínua da base de dados. Um dos ficheiro contém os Canonical SMILES de modo a acelerar a pesquisa por subestrutura, outro contém os fingerprints para acelerar o projecto de pesquisa por similaridade e o outro ficheiro contém os compostos em formato MDL para fornecer ao utilizadores quando este pretender o download da base de dados em formato SDF.

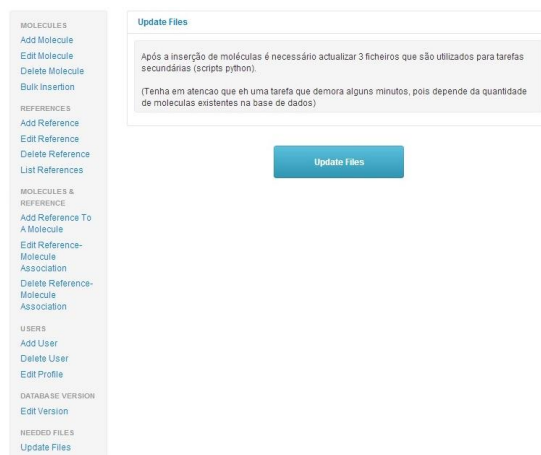


Figura 62 – Actualização de ficheiros

Por motivos de segurança não é possível manter a ligação activa até que o servidor retorne uma resposta de conclusão do processo, então foi necessário realizar uma estimativa da duração da operação baseada na quantidade de compostos existentes na base de dados. E essa informação é transmitida ao administrador (Figura 63).

> O processo de criação dos ficheiros foi iniciado.  
Por favor não adicione, remova ou edite compostos durante os próximos 8 minutos.

Figura 63 – Previsão da duração da actualização dos ficheiros

## 6.6 Transformação e carregamento de dados

Como mencionado anteriormente, os dados necessários para popular a base de dados encontravam-se armazenados e dispersos por várias folhas de cálculo e documentos de texto (Figura 1). Foi desenvolvido um script na linguagem PHP para percorrer os vários ficheiros de cálculo, fazer a correspondência entre as linhas dos mesmos, guardar os dados em variáveis, realizar a sua validação, executar os scripts *Python* sobre as variáveis e ligar-se aos sistemas externos para obtenção de mais dados. Após estes processos os dados foram inseridos na base de dados.

## Capítulo 7

### Conclusão

Um novo modelo de previsão da permeabilidade de um composto na Barreira Hemato-Encefálica encontra-se acessível através da interface web desenvolvida para esse fim e através do web service que permite o acesso ao modelo por parte de aplicações externas, ou seja, os objectivos propostos para a aplicação web B3PP foram atingidos. Os requisitos identificados foram também cumpridos.

A informação sobre compostos cujos valores de permeabilidade já tinham sido validados experimentalmente e que se encontravam armazenados e dispersos em folhas de cálculo encontram-se disponíveis para toda a comunidade através do sistema de informação B3Info. Os mesmos podem ser armazenados num único ficheiro e de forma coerente para fácil utilização. O sistema B3Info disponibiliza várias funcionalidades de pesquisa e funcionalidades de gestão do mesmo e da base de dados. Por fim pode-se concluir que o objectivo de contribuir para a investigação na área da permeabilidade da Barreira Hemato-Encefálica foi atingido e aceite pela comunidade científica pois foi publicado um artigo sobre a aplicação e modelo do B3PP e um *poster* sobre o B3Info.

#### **Desafios encontrados**

No âmbito da permeabilidade na Barreira Hemato-Encefálica foi um grande desafio entender todos os conceitos químicos associados ao tema e perceber como utilizar as tecnologias Python, Pybel e Open Babel para obter dados químicos relacionados com os mesmos.

No aspecto mais técnico o desafio maior foi a integração de vários tipos de linguagens utilizadas e perceber como enviar e receber dados entre as mesmas. Outro desafio foi a definição do layout de resultados de pesquisa pois era necessário a apresentação de muita e variada informação e havia a necessidade de a estruturar da melhor maneira possível.

### **Trabalho futuro**

No âmbito do B3Info seria proveitoso que a base de dados seja populada com mais compostos, referencias e relações composto-referencia.

No âmbito do B3PP seria vantajoso criar uma base de dados que registasse todos os compostos inseridos para predição, para posterior análise. Outra melhoria seria a continua evolução por parte do modelo de predição.





## Capítulo 8

### Bibliografia

[1] Machine Learning Algorithms To Predict Blood-Brain Barrier Permeability Of Drug Molecules, Inês Filipa dos Santos Martins, 2011

[2] Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit, Noel M O’Boile, Chris Morley and Geoffrey R Hutchison

[3] Arthur Dalby, James G. Nourse, W. Douglas Hounshell, Ann K. I.Gushurst, David L. Grier, Burton A. Leland, and John Laufer. Description of several chemical structure file formats used by computer programs developed at molecular design limited. J. Chem. Inf. Comput. Sci., 32(3):244{255, May 1992.

[4] <http://www.r-project.org/>

[5] [7] By Paul Dubois and Pub Date. MySQL Cookbook. Database, (October):1022, 2002.

[6] Weininger, D. *SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules*. J. Chem. Inf. and Comp. Sciences 1988, 28, 31–36.

[7] By David Sklar and Adam Trachtenberg. PHP Cookbook. Number March 2008 in Cookbooks Series. O’Reilly, 2002.

[8] Christopher Schmitt. CSS Cookbook, volume 22. O’Reilly, 2006.

[9] *ThermInfo*: Sistema de Informação para Coligir e Apresentar Propriedades Termoquímicas, Ana Lino Teixaeira, 2009

[10] <http://www.chemspider.com/>

[11] <http://www.drugbank.ca/>

[12] <http://www.ebi.ac.uk/chebi/>

[13] <http://pubchem.ncbi.nlm.nih.gov/>

[14] <http://cactus.nci.nih.gov/>

[15] Frurip, D.; Britton, L.; Fenlon, W.; Going, J; Harrison, B. K.; Niemeier, J.; Ural, E. A. *The Role of ASTM E27 Methods in Hazard Assessment: Part I. Thermal Stability, Compatibility, and Energy Release Estimation Methods*. *Process Saf. Progr.* **2004**, 23, 266–278.

[16] Peter Lubbers, Brian Albers, and Frank Salim. Overview of HTML5. In *Pro HTML5 Programming*, pages 1–23. Apress, 2010.

[17] Linstrom, P.J.; Mallard, W.G. (Eds.) *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg MD, <<http://webbook.nist.gov>> (accessed in May 2009).